



(12) UK Patent (19) GB (11) 2 321 642 (13) B

(54) Title of Invention

Human telomerase reverse transcriptase
promoter

(51) INT CL⁷: C12N 9/12, A61K 38/46, A61P 35/00, C07K
16/18 16/40, C12N 15/54, C12Q 1/34 1/68

(21) Application No
9804859.8

(22) Date of filing
01.10.1997

(22) Date Lodged
06.03.1998

(30) Priority Data

(31) 08724643
08844419
08846017
08851843
08854050
08911312
08912951
08915503

(32) 01.10.1996
18.04.1997
25.04.1997
06.05.1997
09.05.1997
14.08.1997
14.08.1997
14.08.1997

(33) US

(62) Divided from Application No.
9720890.4 under Section
15(4) of the Patents Act 1977

(43) Application published
05.08.1998

(45) Patent published
09.02.2000

(52) Domestic classification
(Edition R)
C3H HB7E H645 H650 H651
H652 H653 H654 H655 H656
H657 H720

(72) Inventor(s)
Thomas R Cech
Joachim Lingner
Toru Nakamura
Karen B Chapman
Gregg B Morin
Calvin B Harley
William H Andrews

(73) Proprietor(s)
Geron Corporation

(Incorporated in USA -
California)

230 Constitution Drive
Menlo Park
California 94025
United States of America

University Technology
Corporation

(Incorporated in USA -
Colorado)

Suite 250
3101 Iris Avenue
Boulder
Colorado 80301
United States of America

(74) Agent and/or
Address for Service
Hepworth Lawrence Bryer
& Bizley
Merlin House
Falconry Court
Bakers Lane
Epping
Ess x
CM16 5DQ
Unit d Kingd m

(continued n n xt pag)

GB 2 321 642 B - continuation

- (52) Domestic classification
(Edition R) (cont'd)
C6Y YB YF YG08 YG09 YM
Y115 Y125 Y130 Y134 Y156
Y162 Y189 Y192 Y313 Y319
Y327 Y330 Y331 Y332 Y333
Y338 Y405 Y406 Y407 Y410
Y420 Y501 Y503
U1S S1068 S1284 S1285
S1289 S1303 S1313 S1332
S1333 S1334 S1337 S2411
S2415 S2416 S2417 S2419
- (56) Documents cited
WO96/01835 A1
Science 1997, 276, 561-567
Science 1997, 277, 955-959
Trends in Cell Biology 1997,
7, 299-302
Cell 1997, 90, 785-795
Current Opinion in Cell
Biology 1996, 8, 374-380
J. Biological Chemistry 1995,
270(42), 24623-24626
Science 1995, 269, 1236-1241
- (58) Field of search

As for published application
2321642 A viz:
UK CL(Edition P) **C3H HB7E**
HB7M
INT CL⁶ C12N 9/12
Online: WPI, CLAIMS,
DIALOG/BIOTECH
updated as appropriate

	human	tez1	EST2	p123	Motif 0
					AKFLHLMMSVYVVELLSRFFVYTETTFQKNR
					ISEIEMVLGKRSNAKMCLSDFEKRQIFAEFIYWLNSFIIPILQSFYFITESSDLNR
					LKDFRWLFISD---IWFTKHNFENLQLAICFI SWLFRQLIPKIIQTFFCYCTEISSTVT-
					TREISNMQVET-SAKHFYFFDHEN-IYVLWKLRLWIFEDLWVSLRRCFFVYVTEQQKSYSK
					+++ + + + + + + +

human
tez1
EST2
p123

Motif 2

human	RPVNM DYVVGARTFRREKRAERLTSRVKALF-SVLNYERA	
tez1	RLITN-LRKRF LKMGSNKKMLVSTNQTLRPVASILKHLIN EESSGIPFNLEVMKLLTF	
EST2	RIIAIPCRGADEEFTIYKHNHKNAIQPTQKILEYLRNKRPTSTFKIYSPTQIADRIKEF	
p123	RPIMTFNKKIVNSDRKTTKLTNTTKLLNSHLM LKTLKN-RMFKDPFGFAVFNYYDDVMKKY	* *

tez1
EST2
p123

FIG. 1

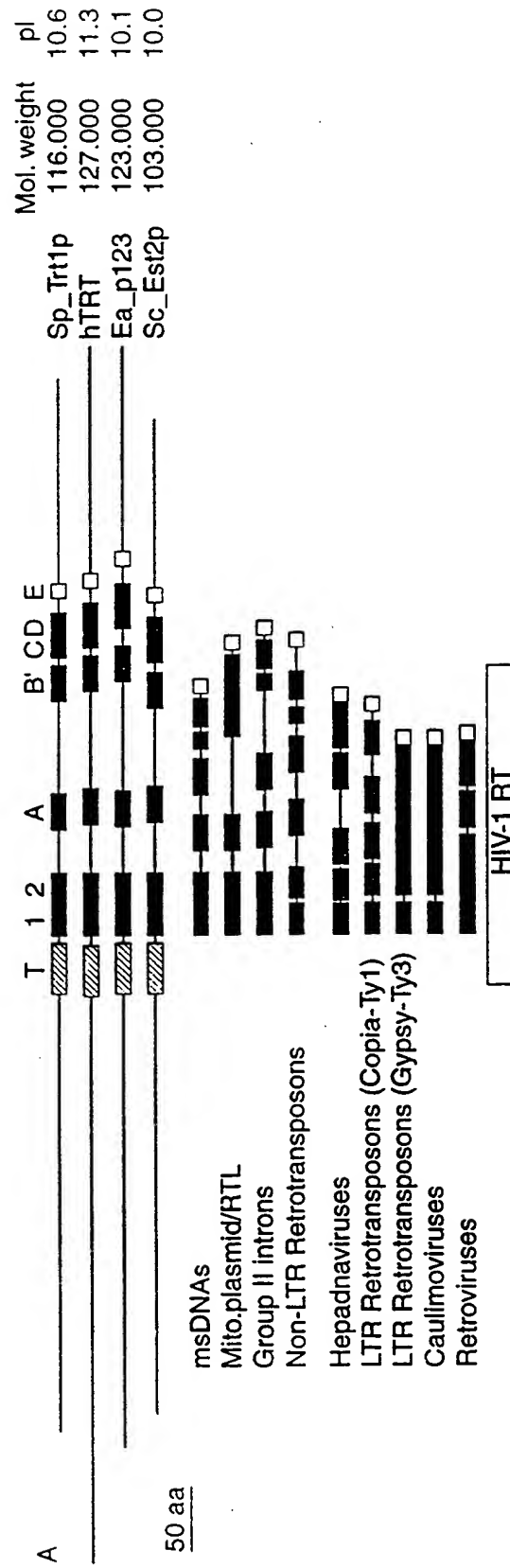


FIG. 2

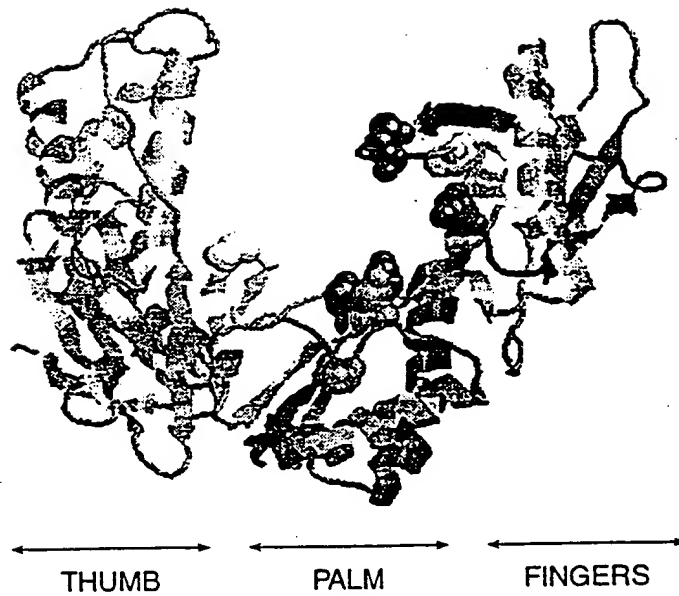


FIG. 3

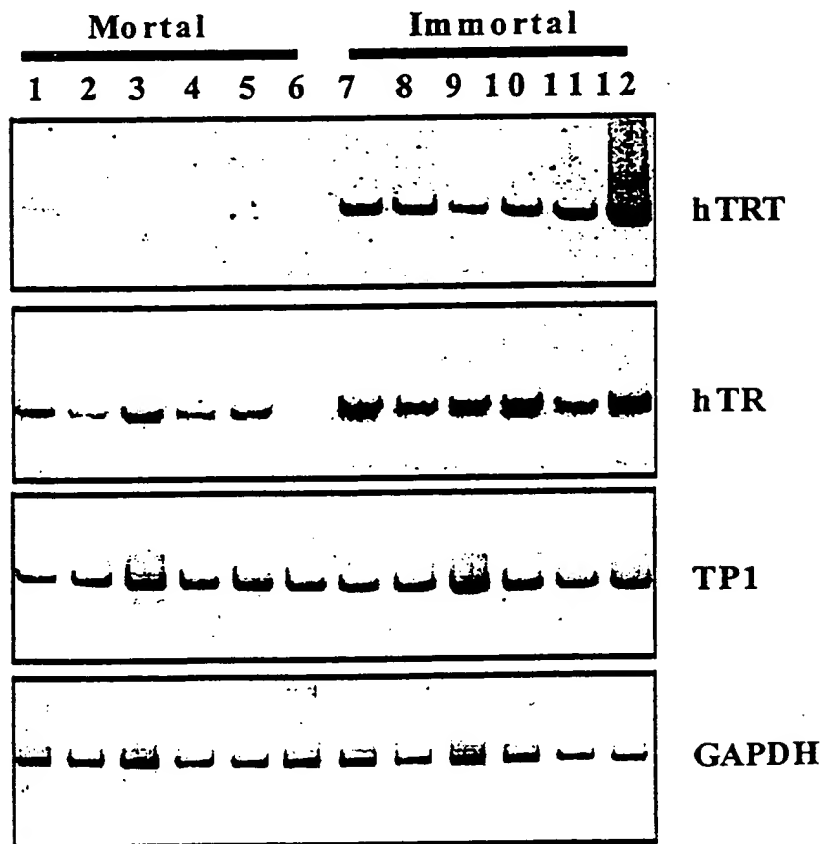


FIG. 5

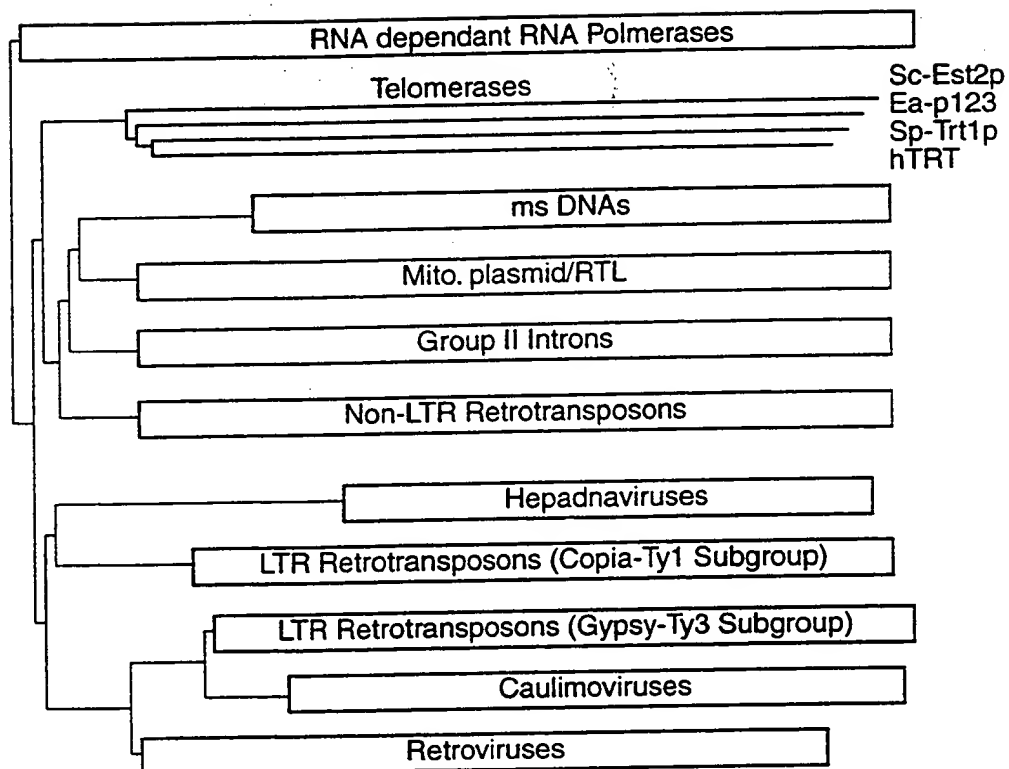


FIG. 6

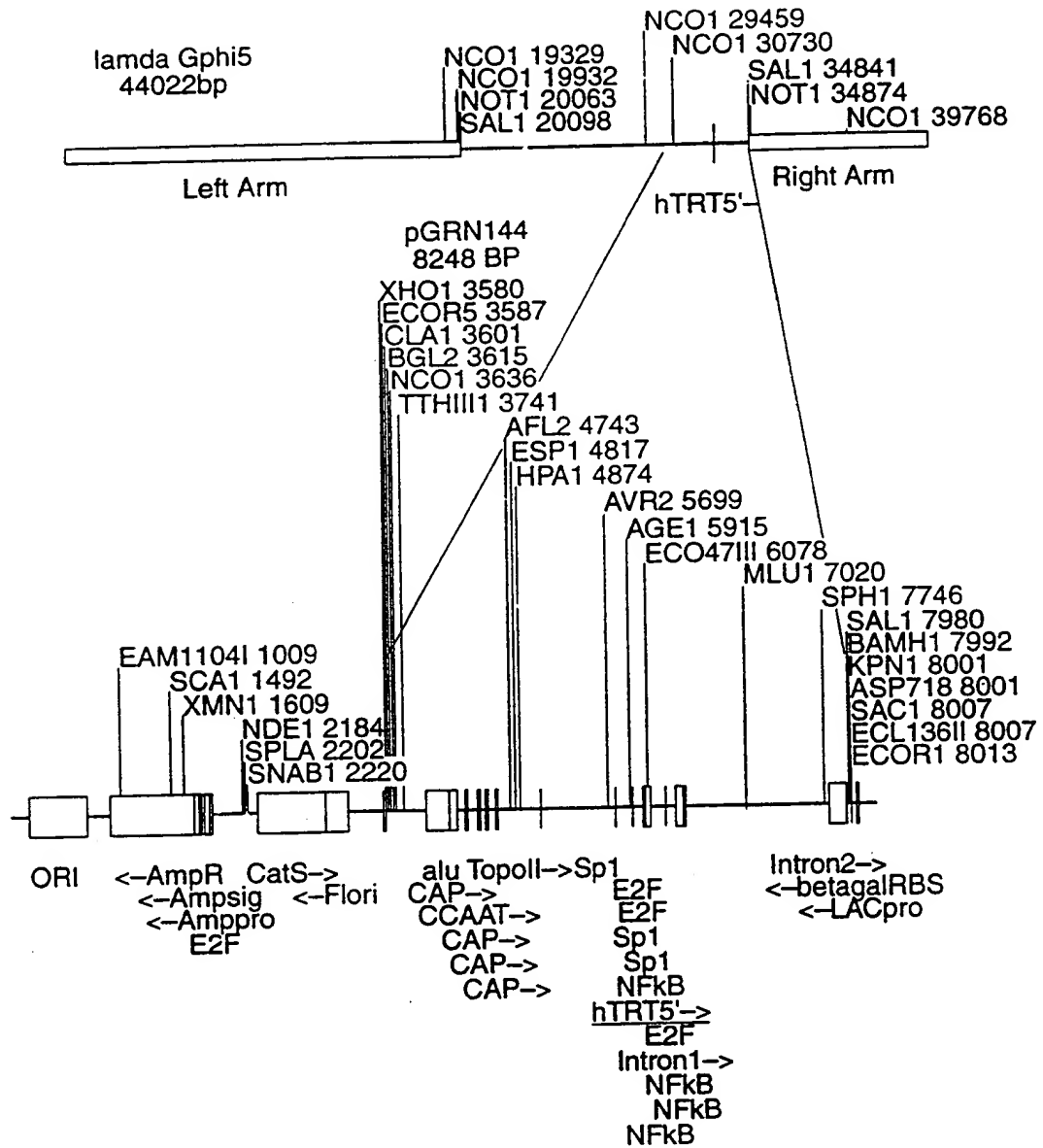


FIG. 7

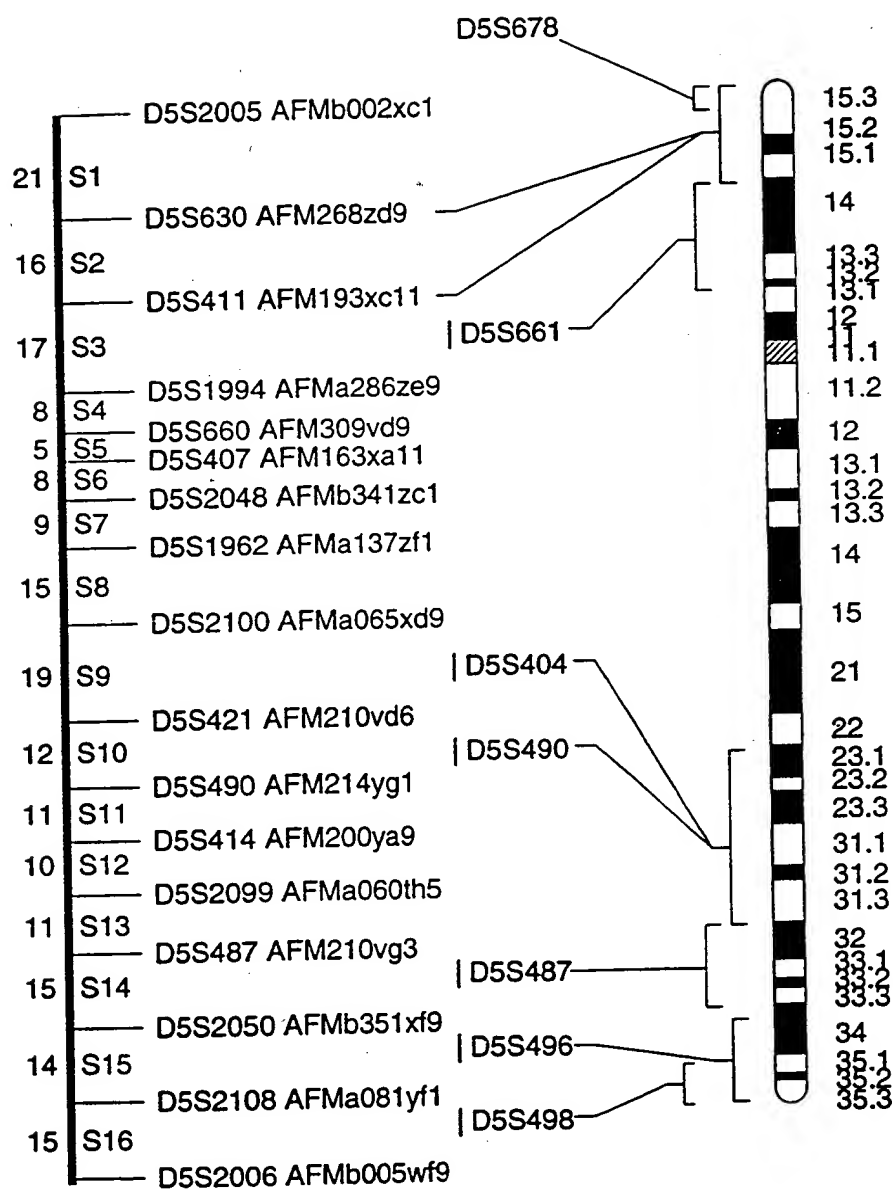


FIG. 8

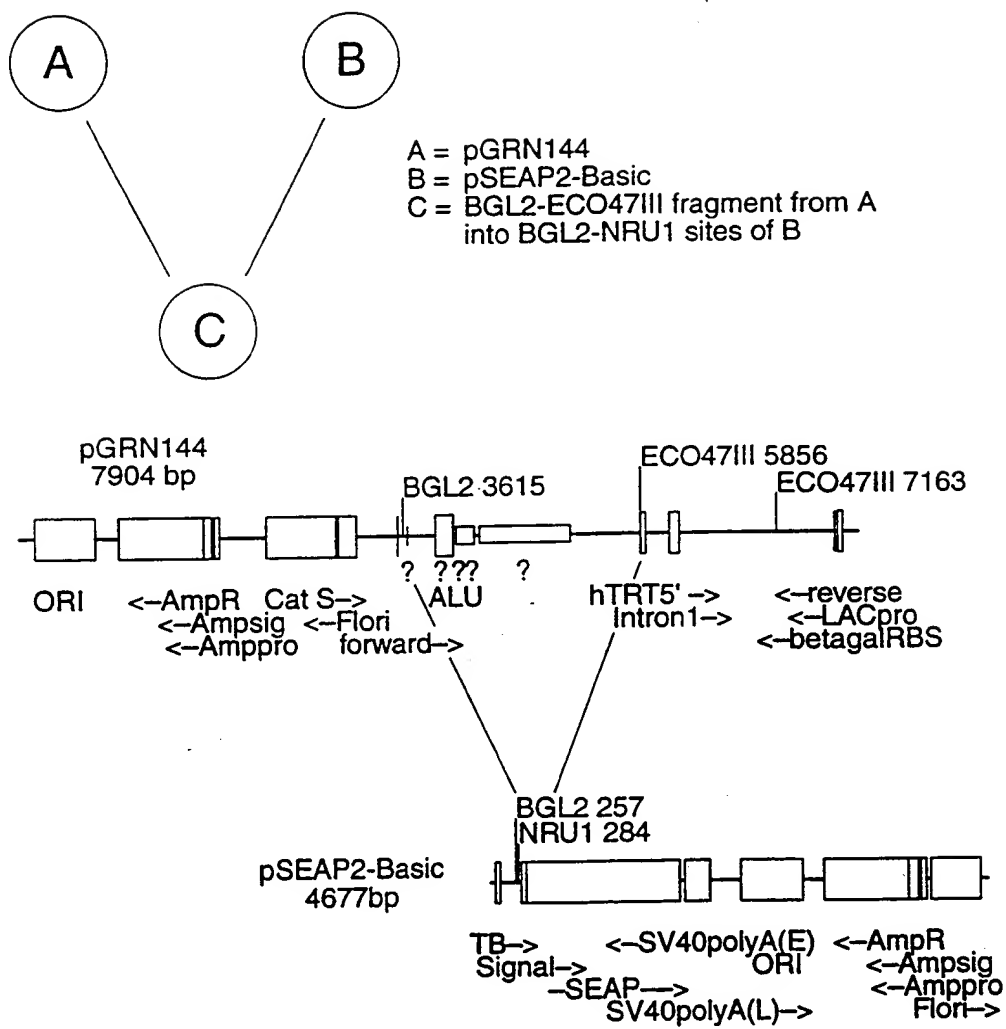


FIG. 9

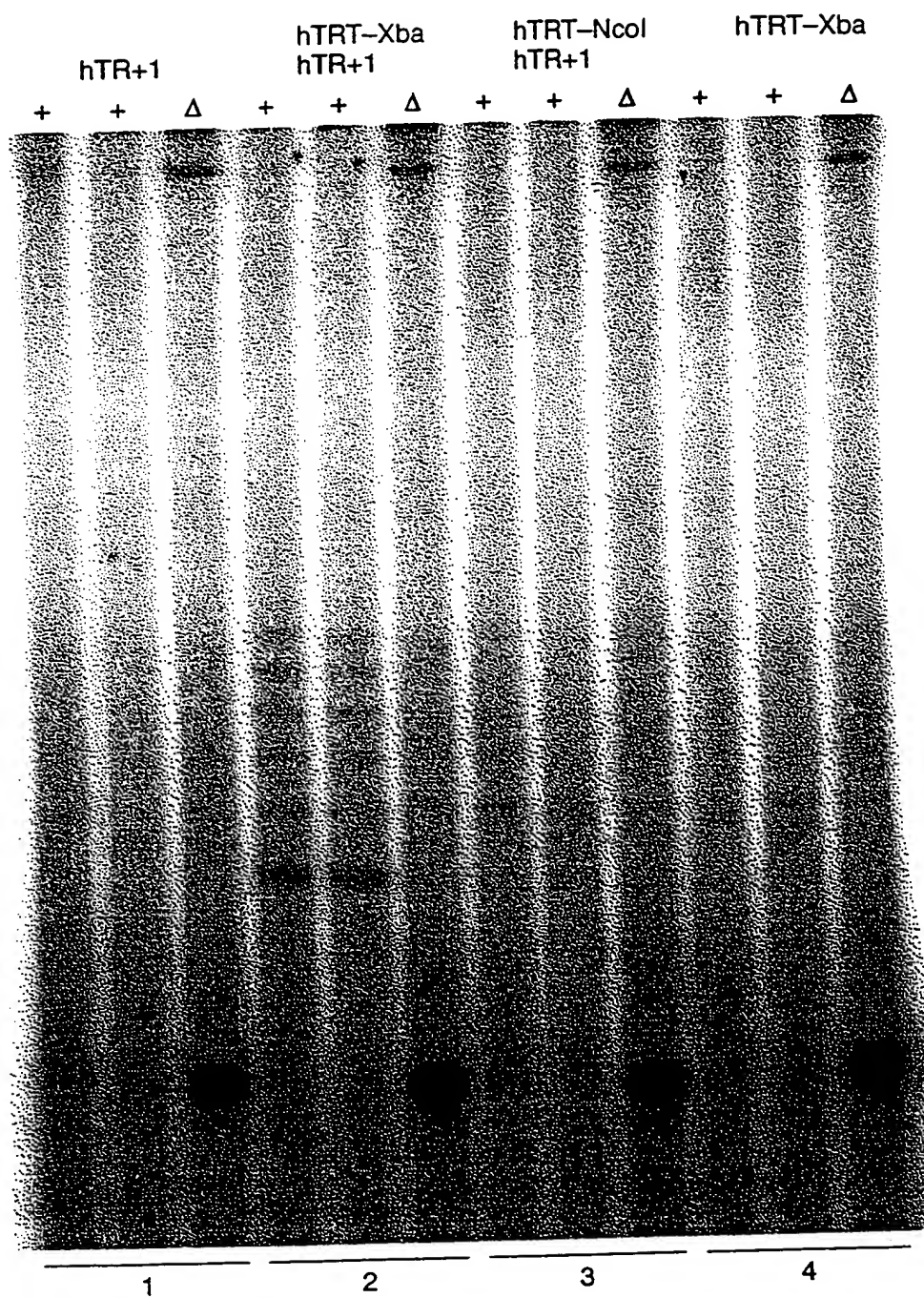


FIG. 10A

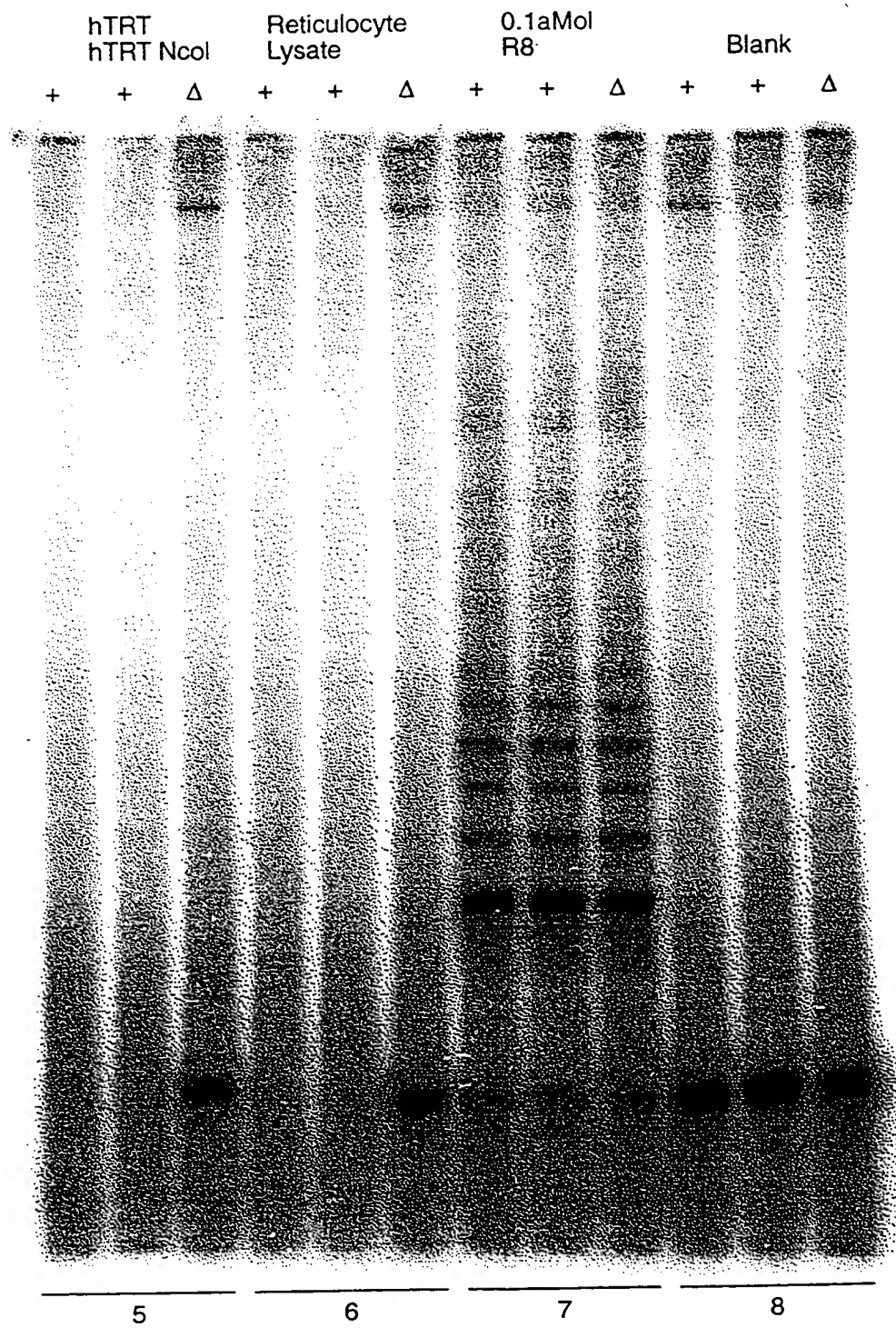


FIG. 10B

Telomerase Specific Motifs

	MOTIF T	MOTIF T'
TRT con	W1	
hTRT	546 WLMSVYVVELLRSEFFVVTETTFQKNRLFFYRKSVWSKLQSIGI 13 EAEVR	E V
spTRT	429 WLXNSFIIPILQSFYITESSDLRNRVTYFRKDIWKLRCRPF 12 ENNVR	
Ea_p123	441 WIFEDLVVSLIRCFYVTEQQKSYSKTYYYRKNIDVIMKMSI 12 EKEVE	
Sc_Est2	366 WLFRLIPKIIQTFYCYCTEISSTVT.IVYFRHDTWKLITPFI 9 ENNVC	

	MOTIF 1	MOTIF 2	MOTIF A	MOTIF B'
TRT con	R ipkk	fr i	p lyf D cyd i	Y q GIPQGs 1S 1 Y
hTRT	11 SRLRFIPKPDG 0 LRPIV		69 PELYFVKVDVTGAYDTI 104 YVQCQGIPOGSILSTLLCSLCY	
spTRT	10 AVIRLLPKKNT 0 FRLIT		66 RKKYFVRIDIKSCYDRI 99 YLQKVGIPQGSILSSFLCHFYM	
Ea_p123	10 GKRLIPKKT 0 FRPIM		67 PKLFFATMDIEKCYDSV 117 YKQTKGIPQGLCVSSILSSFY	
Sc_Est2	13 SKMRIIPKKS 2 FRIIA		68 PELYFMKFDVKSCYDSI 85 YIREDGLFQGSLSAPIVDLVY	
RT con	p hh h k	hr h	h hdh AF h GY	hpQG pp hh h

Telomerase RT Motifs (Fingers)

Telomerase RT Motifs (Palm, Primer Grip)

	MOTIF C	MOTIF D	MOTIF E
TRT con	111rl DDfL it	g n K	w g s 1
hTRT	15 LLLRLVDDFLLVT 15 GVPEYGCVVNLKRTVV 24 WCGLLDTRTL		192
spTRT	16 VLLRVVDDFLFIT 15 GFEKHNFSTSLKTVI 22 FFGFSVNMRS		176
Ea_p123	24 LLMRLTDDYLLIT 15 VSRENGEFKNKKLQT 28 WIGISIDMKTL		174
Sc_Est2	18 LILKLADDFLIIS 15 GFQKYNKANRDKILA 25 WKHSSTMNHF		141
RT con	h Y DDhhh	Gh h ck h hLG h	

FIG. 11

181 GGACCCGGCGGCTTTCCGCGCGCTGGTGGCCCAGTGCCTGGTGTGCGTGCCCTGGGACGC
CCTGGGCCGCCGAAAGGCGCGGACCACCGGGTCACGGACCACACGCACGGGACCCTGCG

NFkB_CS1
GGGRQTYYYQC
NFkB-MHC-I.2
TGGGCTTCCCC

241 ACGGCCGCCCCCGCCGCCCCCTCCTTCCGCCAGGTGGGCCTCCCCGGGGTCGGCGTCCG
TGCCGGCGGGGGCGGCGGGGAGGAAGGCGGTCCACCCGGAGGGCCCCAGCCGCAGGC

Intron1

301 GCTGGGGTTGAGGGCGGCCGGGGGGAACCAGCGACATGCGGAGAGCAGCGCAGGCGACTC
CGACCCCAACTCCCGCCGGCCCCCCTTGGTCGCTGTACGCCTCTCGTCGCGTCCGCTGAG

NFkB_CS1
GGGRQTYYYQC
NFkB_CS2
RGGGRMTYYCC
Topo_II_cleavage_site
RNYNNCNGYNGKTNINY
*****>

361 AGGGCGCTTCCCCCGCAGGTGTCCTGCCTGAAGGAGCTGGTGGCCCAGTGCTGCAGAGG
TCCCGCGAAGGGGGCGTCCACAGGACGGACTTCCTCGACCACCGGGCTCACGACGTCTCC

FIG. 12

1	AAAACCCCAA	AACCCCAAAA	CCCCTTTT	AGCCCTGCAG	TTGGAAATAT
51	AACCTCAGTA	TTAATAAGCT	CAGATTTTAA	ATATTAATTA	CAAAACCTAA
101	ATGGAGGTTG	ATGTTGATAA	TCAAGCTGAT	AATCATGGCA	TTCACCTCAGC
151	TCTTAAGACT	TGTGAAGAAA	TTAAAGAAGC	TAAAACGTTG	TACTCTTGGA
201	TCCAGAAAGT	TATTAGATGA	AGAAATCAAT	CTCAAAGTCA	TTATAAAGAT
251	TTAGAAGATA	TTAAAATATT	TGCGCAGACA	AATATTGTTG	CTACTCCACG
301	AGACTATAAT	GAAGAAGATT	TTAAAGTTAT	TGCAAGAAAA	GAAGTATTTT
351	CAACTGGACT	AATGATCGAA	CTTATTGACA	AATGCTTAGT	TGAACTTCTT
401	TCATCAAGCG	ATGTTTCAGA	TAGACAAAAA	CTTCAATGAT	TTGGATTTC
451	ACTTAAGGGA	AATCAATTAG	CAAAGACCCA	TTTATTAACA	GCTCTTTCAA
501	CTCAAAAAGCA	GTATTTCTTT	CAAGACGAAT	GGAACCAAGT	TAGAGCAATG
551	ATTGGAAATG	AGCTCTTCCG	ACATCTCTAC	ACTAAATATT	TAATATTCCA
601	GCGAACTTCT	GAAGGAACTC	TTGTTCAATT	TTGCGGGAAT	AACGTTTTTG
651	ATCATTTGAA	AGTCAACGAT	AAGTTTGACA	AAAAGCAAAA	AGGTGGAGCA
701	GCAGACATGA	ATGAACCTCG	ATGTTGATCA	ACCTGCAAT	ACAATGTCAA
751	GAATGAGAAA	GATCACTTTC	TCAACAACAT	CAACGTGCCG	AATTGGAATA
801	ATATGAAATC	AAGAACCAGA	ATATTTTATT	GCACTCATTT	TAATAGAAAT
851	AACCAATTCT	TCAAAAAGCA	TGAGTTTGTG	AGTAACAAAA	ACAATATTTT
901	AGCGATGGAC	AGAGCTCAGA	CGATATTCAC	GAATATATTC	AGATTTAATA
951	GAATTAGAAA	GAAGCTAAAA	GATAAGGTTA	TCGAAAAAAT	TGCCTACATG
1001	CTTGAGAAAG	TCAAAGATTT	TAACCTCAAC	TACTATTTAA	CAAAATCTTG
1051	TCCTCTTCCA	GAAAATTGGC	GGGAACGGAA	ACAAAAAATC	GAAAACCTGA
1101	TAAATAAAAC	TAGAGAAGAA	AAGTCGAAGT	ACTATGAAGA	GCTGTTTAGC
1151	TACACAACCTG	ATAATAAATG	CGTCACACAA	TTTATTAATG	AATTTTCTTA
1201	CAATATACTC	CCCAAAGACT	TTTGTACTGG	AAGAAACCGT	AAGAATTTTC
1251	AAAAGAAAGT	TAAGAAATAT	GTGGAACATA	ACAAGCATGA	ACTCATTCAC
1301	AAAAAATTAT	TGCTTGAGAA	GATCAATACA	AGAGAAATAT	CATGGATGCA
1351	GGTTGAGACC	TCTGCAAAGC	ATTTTTATTA	TTTTGATCAC	GAAAACATCT
1401	ACGCTTTATG	GAAATTGCTC	CGATGGATAT	TCGAGGATCT	CGTCGCTCG
1451	CTGATTAGAT	GATTTTTCTA	TGTCACCGAG	CAACAGAAAA	GTTACTCCAA
1501	AACCTATTAC	TACAGAAAGA	ATATTTGGGA	CGTCATTATG	AAAATGTCAA
1551	TCGCAGACTT	AAAGAAGGAA	ACGCTTGCTG	AGGTCCAAGA	AAAAGAGGTT
1601	GAAGAATGGA	AAAAGTCGCT	TGGATTTGCA	CCTGGAAAAAC	TCAGACTAAT
1651	ACCGAAGAAA	ACTACTTTCC	GTCCAATTAT	GACTTTCAAT	AAGAAGATTG
1701	TAAATTCAGA	CCGGAAGACT	ACAAAATTAA	CTACAAATAC	GAAGTTATTG
1751	AACTCTCACT	TAATGCTTAA	GACATTGAAG	AATAGAATGT	TTAAAGATCC
1801	TTTTGGATTG	GCTGTTTTTA	ACTATGATGA	TGTAATGAAA	AAGTATGAGG
1851	AGTTTGTGTTG	CAAATGGAAG	CAAGTTGGAC	AACCAAAACT	CTTCTTTGCA
1901	ACTATGGATA	TCGAAAAGTG	ATATGATAGT	GTAAACAGAG	AAAAACTATC
1951	AACATTCCCTA	AAAACACTTA	AATTACTTTC	TTCAAGATTTT	TGGATTATGA
2001	CTGCACAAAT	TCTAAAGAGA	AAGAATAACA	TAGTTATCGA	TTTCGAAAAAC
2051	TTTAGAAAGA	AAGAAATGAA	AGATTATTTT	AGACAGAAAT	TCCAGAAGAT
2101	TGCACCTGAA	GGAGGACAAT	ATCCAACCTT	ATTCAGTGTT	CTTGAAAATG
2151	AACAAAATGA	CTTAAATGCA	AAGAAAACAT	TAATTGTTGA	AGCAAAGCAA
2201	AGAAATTATT	TTAAGAAAGA	TAACCTACTT	CAACCAGTCA	TTAATATTTG
2251	CCAATATAAT	TACATTAAC	TTAATGGGAA	GTTTTATAAA	CAACAAAAG
2301	GAATTCCTCA	AGGTCTTTGA	GTTTCATCAA	TTTGTGCATC	ATTTTATTAT
2351	GCAACATTAG	AGGAAAGCTC	CTTAGGATTC	CTTAGAGATG	AATCAATGAA

FIG. 13

```

2401 CCCTGAAAAT CCAAATGTTA ATCTTCTAAT GAGACTTACA GATGACTATC
2451 TTTTGATTAC AACTCAAGAG AATAATGCAG TATTGTTTAT TGAGAAACTT
2501 ATAAACGTAA GTCGTGAAAA TGGATTTAAA TTCAATATGA AGAAACTACA
2551 GACTAGTTTT CCATTAAGTC CAAGCAAATT TGCAAAATAC GGAATGGATA
2601 GTGTTGAGGA GCAAAATATT GTTCAAGATT ACTGCGATTG GATTGGCATC
2651 TCAATTGATA TGAAAACCTCT TGCTTTAATG CCAAATATTA ACTTGAGAAT
2701 AGAAGGAATT CTGTGTACAC TCAATCTAAA CATGCAAACA AAGAAAGCAT
2751 CAATGTGGCT CAAGAAGAAA CTAAAGTCGT TTTTAATGAA TAACATTACC
2801 CATTATTTTA GAAAGACGAT TACAACCGAA GACTTTGCGA ATAAAACTCT
2851 CAACAAGTTA TTTATATCAG GCGGTTACAA ATACATGCAA TGAGCCAAAG
2901 AATACAAGGA CCACTTTAAG AAGAACTTAG CTATGAGCAG TATGATCGAC
2951 TTAGAGGTAT CTAAAATTAT ATACTCTGTA ACCAGAGCAT TCTTTAAATA
3001 CCTTGTGTGC AATATTAAGG ATACAATTTT TGGAGAGGAG CATTATCCAG
3051 ACTTTTTCCT TAGCACACTG AAGCACTTTA TTGAAATATT CAGCACAAAA
3101 AAGTACATTT TCAACAGAGT TTGCATGATC CTCAAGGCAA AAGAAGCAAA
3151 GCTAAAAAGT GACCAATGTC AATCTCTAAT TCAATATGAT GCATAGTCGA
3201 CTATTCTAAC TTATTTTGGG AAGTTAATTT TCAATTTTGG TCTTATATAC
3251 TGGGGTTTTG GGGTTTTGGG GTTTTGGGG

```

FIG. 13
(CONTINUED)

```

1 MEVDVDNQAD NHGIHSALKT CEEIKEAKTL YSWIQKVIRC RNQSQSHYKD
51 LEDIKIFAQT NIVATPRDYN EEDFKVIARK EVFSTGLMIE LIDKCLVELL
101 SSSDVSDRQK LQCFGFQKLG NQLAKTHLLT ALSTQKQYFF QDEWNQVRAM
151 IGNELFRHLY TKYLIFQRTS EGTLVQFCGN NVFDHLKVND KFDKKQKGGG
201 ADMNEPRCCS TCKYNVKNEK DHFLNNINVP NWNMMKSRTT IFYCTHFNRN
251 NQFFKKHEFV SNKNNISAMD RAQTIFTNIF RFNRIKKLK DKVIEKIAYM
301 LEKVKDFNFN YYLTKSCPLP ENWRERKQKI ENLINKTREE KSKYYEELFS
351 YTTDNKCVTQ FINEFFYNIL PKDFLTGRNR KNFQKKVKKY VELNKHELIIH
401 KNLLLEKINT REISWMQVET SAKHFYYFDH ENIYVLWKLL RWIFEDLVVS
451 LIRCFYVTE QOKSYSKTY YRKNIWVIM KMSIADLKKE TLAEVQKEV
501 EEWKKS LGFA PGKLRLIPKK TTFRPIMTFN KKI VNSDRKT TKLTNTTKLL
551 NSHMLKTLK NRMFKDPFGF AVFNYDDVMK KYEEFVCKWK QVGQPKLFFA
601 TMDIEKCYDS VNREKLSTFL KTTKLLSSDF WIMTAQILKR KNNIVIDSKN
651 FRKKEMKD YF RQKFQKIALE GGQYPTLFSV LENEQNDLNA KKT LIVEAKQ
701 RNYFKKDNLL QPVINICQYN YINFNGKFYK QTKGIPQGLC VSSILSSFY Y
751 ATLEESSLGF LRDESMNPEN PNVNLLMRLT DDYLLITTQE NNAVL FIEKL
801 INVSRENGFK FNMKKLQTSF PLSPSKFAKY GMDSV EEQNI VQDYCDWIGI
851 SIDMKTALM PNINLRIE GI LCTLNLMQT KKASMWLKKK LKSFLMNNIT
901 HYFRKTITTE DFANKTLNKL FISGGYKYMQ CAKEYKDHFK KNLAMSSMID
951 LEVSKIIYSV TRAFFKYLVC NIKDTIFGEE HYPDFFLSTL KHFIEIFSTK
1001 KYIFNRVCM I LKAKEAKLKS DQCQSLIQYD A

```

FIG. 14

1 ggtaaccgatttactttccttcttcataagaagtaattgcttctcctcgaacgctcctaaatctcttggaataatttttacaaga 80
 81 actcaataacaataaccagaagtcattccaatatgaaggtgtatttagtgatcgataataatttcttatttctatcggtcgctta 160
 161 ccaagtataaggacaaaaaagaacaaacttctccctcaagaactttactttatttaatttacttttcaaatatttcg 240
 241 ggttcggttacttttaactcgttggtactgttttagtgctacttagcgaacccggtgtttctaccgcgtcattggatat 320
 321 agtcttggagtagctcacagaaaatccttcaaaaatccttgatgagactatattagactcattacacgctgcataatc 400
 401 ttaacatggagccttacacttttagatgagtcgcatgagtgatatttggtatccatccacggttgccttgaaaaag 480
 481 gttgataatttgcataaattctccttagtggtgtaataccgcgaagtttttttgatgcttgcacacgcttagcatg 560
 561 attgagatattcaaaaatttctatccactacaactcctttaacgcggttttatttttctatttctcattcattgtgtt 640
 641 ccaaatatgtatcctcgtattaggcttttttccggttttactcctcctggaatcgtaacctttttcactattccccctaaatga 720
 721 ataactcaaaattagtttcgcttataattgtagtagtagaagaattggtgattctactcgtgtaaatgttattagtttaa 800
 801 gatactttgcaaaaacattttagctatcattatataaaaaaaatccataaattataaaatcaatatttgcggtc 880
 881 actatttttaaaacgtttatgatcagtaggacacttttgcatatatatagttatgcttaattggttacttgaacttgc 958

 959 ATG ACC GAA CAC CAT ACC CCC AAA AGC AGG ATT CTT CGC TTT CTA GAG AAT CAA TAT GTA 1018
 1 M T E H H T P K S R I L R F L E N Q Y V 20

 1019 TAC CTA TGT ACC TTA AAT GAT TAT GTA CAA CTT GTT TTG AGA GGG TCG CCG GCA AGC TCG 1078
 21 Y L C T L N D Y V Q L V L R G S P A S S 40

 1079 TAT AGC AAT ATA TGC GAA CGC TTG AGA AGC GAT GTA CAA ACG TCC TTT TCT ATT TTT CTT 1138
 41 Y S N I C E R L R S D V Q T S F S I F L 60

 1139 CAT TCG ACT GTA GTC GGC TTC GAC AGT AAG CCA GAT GAA GGT GTT CAA TTT TCT TCT CCA 1198
 61 H S T V V G F D S K P D E G V Q F S S P 80

 1199 AAA TGC TCA CAG TCA GAG gtatatataatttttggatttttttctattcgggtagctaatatatatgggcag 1272
 81 K C S Q S E 86

 1273 CTA ATA GCG AAT GTT GTA AAA CAG ATG TTC GAT GAA AGT TTT GAG CGT CGA AGG AAT CTA 1332
 87 L I A N V V K Q M F D E S F E R R N L 106

 1333 CTG ATG AAA GGG TTT TCC ATG gtaaggatttcttaattgtgaaatatttaccctgcaattactgtttcaagaga 1405
 107 L M K G F S M 113

 1406 ttgtatttaaccgataaag AAT CAT GAA GAT TTT CGA GCC ATG CAT GTA AAC GGA GTA CAA AAT 1469
 114 N H E D F R A M H V N G V Q N 128

FIG. 15

1470 GAT CTC GTT TCT ACT TTT CCT AAT TAC CTT ATA TCT ATA CTT GAG TCA AAA AAT TGG CAA 1529
 129 D L V S T F P N Y L I S I L E S K N W Q 148
 1530 CTT TTG TTA GAA AT gtaataaccgggtaagatgttgccgacacttgaacaagactgacaagtatag T ATC GGC 1601
 149 L L L E I 155
 1602 AGT GAT GCC ATG CAT TAC TTA TTA TCC AAA GGA AGT ATT TTT GAG GCT CTT CCA AAT GAC 1661
 156 S D A M H Y L L S K G S I F E A L P N D 175
 1662 AAT TAC CTT CAG ATT TCT GGC ATA CCA CTT TTT AAA AAT AAT GTG TTT GAG GAA ACT GTG 1721
 176 N Y L Q I S G I P L F K N N V F E E T V 195
 1722 TCA AAA AAA AGA AAG CGA ACC ATT GAA ACA TCC ATT ACT CAA AAT AAA AGC GCC CGC AAA 1781
 196 S K K R K R T I E T S I T Q N K S A R K 215
 1782 GAA GTT TCC TGG AAT AGC ATT TCA ATT AGT AGG TTT AGC ATT TTT TAC AGG TCA TCC TAT 1841
 216 E V S W N S I S I S R F S I F Y R S S Y 235
 1842 AAG AAG TTT AAG CAA G gtaactaatactgttatccttcataactaatttttag AT CTA TAT TTT AAC 1907
 236 K K F K Q D L Y F N 245
 1908 TTA CAC TCT ATT TGT GAT CGG AAC ACA GTA CAC ATG TGG CTT CAA TGG ATT TTT CCA AGG 1967
 246 L H S I C D R N T V H M W L Q W I F P R 265
 1968 CAA TTT GGA CTT ATA AAC GCA TTIT CAA GTG AAG CAA TTG CAC AAA GTG ATT CCA CTG GTA 2027
 266 Q F G L I N A F Q V K Q L H K V I P L V 285
 2028 TCA CAG AGT ACA GTT GTG CCC AAA CGT CTC CTA AAG GTA TAC CCT TTA ATT GAA CAA ACA 2087
 286 S Q S T V V P K R L L K V Y P L I E Q T 305
 2088 GCA AAG CGA CTC CAT CGT ATT TCT CTA TCA AAA GTT TAC AAC CAT TAT TGC CCA TAT ATT 2147
 306 A K R L H R I S L S K V Y N H Y C P Y I 325
 2148 GAC ACC CAC GAT GAT GAA AAA ATC CTT AGT TAT TCC TTA AAG CCG AAC CAG GTG TTT GCG 2207
 326 D T H D E K I L S Y S L K P N Q V F A 345
 2208 TTT CTT CGA TCC ATT CTT GTT CGA GTG TTT CCT AAA TTA ATC TGG GGT AAC CAA AGG ATA 2267
 346 F L R S I L V R V F P K L I W G N Q R I 365

FIG. 15
(CONTINUED)

2268 TTT GAG ATA ATA TTA AAA G gttattgtataaaattttattaccactaacgattttaccag AC CTC GAA ACT 2336
 366 F E I I L K D L E T 375
 2337 TTG AAA TTA TCG AGA TAC GAG TCT TTT AGT TTA CAT TAT TTA ATG AGT AAC ATA AAG 2396
 376 F L K L S R Y E S F S L H Y L M S N I K 395
 2397 gtaatatgccaatttttttaccatttaataacaatcag ATT TCA GAA ATT GAA TGG CTA GTC CTT GGA 2465
 396 I S E I E W L V L G 405
 2466 AAA AGG TCA AAT GCG AAA ATG TGC TTA AGT GAT TTT GAG AAA CGC AAG CAA ATA TTT GCG 2525
 406 K R S N A K M C L S D F E K R K Q I F A 425
 2526 GAA TTC ATC TAC TGG CTA TAC AAT TCG TTT ATA ATA CCT ATT TTA CAA TCT TTT TTT TAT 2585
 426 E F I Y W L Y N S F I I P I L Q S F F Y 445
 2586 ATC ACT GAA TCA AGT GAT TTA CGA AAT CGA ACT GTT TAT TTT AGA AAA GAT ATT TGG AAA 2645
 446 I T E S S D L R N R T V Y F R K D I W K 465
 2646 CTC TTG TGC CGA CCC TTT ATT ACA TCA ATG AAA GCG TTT GAA AAA ATA AAC GAG 2705
 466 L C R P F I T S M K M E A F E K I N E 485
 2706 gtattttaagttattttttgcaaaaagctaatttttcag AAC AAT GTT AGG ATG GAT ACT CAG AAA ACT 2775
 486 N N V R M D T Q K T 495
 2776 ACT TTG CCT CCA GCA GTT ATT CGT CTA TTA CCT AAG AAG AAT ACC TTT CGT CTC ATT ACG 2835
 496 T L P P A V I R L L P K K N T F R L I T 515
 2836 AAT TTA AGA AAA AGA TTC TTA ATA AAG gtatttaatttttggtcatcaatgtactttacttctaatctatta 2906
 516 N L R K R F L I K 524
 2907 ttagcag ATG GGT TCA AAC AAA ATG TTA GTC AGT ACG AAC CAA ACT TTA CGA CCT GTG 2967
 525 M G S N K K M L V S T N Q T L R P V 542
 2968 GCA TCG ATA CTG AAA CAT TTA ATC AAT GAA GAA AGT AGT GGT ATT CCA TTT AAC TTG GAG 3027
 543 A S I L K H L I N E S S G I P F N L E 562
 3028 GTT TAC ATG AAG CTT CTT ACT TTT AAG AAG GAT CTT CTT AAG CAC CGA ATG TTT GG gtaat 3088
 563 V Y M K L L T F K K D L L K H R M F G 581

FIG. 15
(CONTINUED)

3089 tatataatgcgagcattcattatttaattttgcag G CGT AAG AAG TAT TTT GTA CGG ATA GAT ATA 3155
 582 R K K Y F V R I D I 591
 3156 AAA TCC TGT TAT GAT CGA ATA AAG CAA GAT TTG ATG TTT CGG ATT GTT AAA AAG AAA CTC 3215
 592 K S C Y D R I K Q D L M F R I V K K L 611
 3216 AAG GAT CCC GAA TTT GTA ATT CGA AAG TAT GCA ACC ATA CAT GCA ACA AGT GAC CGA GCT 3275
 612 K D P E F V I R K Y A T I H A T S D R A 631
 3276 ACA AAA AAC TTT GTT AGT GAG GCG TTT TCC TAT T gtaagttattttttcattggaattttttaacaa 3343
 632 T K N F V S E A F S Y F 643
 3344 attcttttttag TT GAT ATG GTG CCT TTT GAA AAA GTC GTG CAG TTA CTT TCT ATG AAA ACA 3405
 644 D M V P F E K V V Q L L S M K T 659
 3406 TCA GAT ACT TTG TTT GTT GAT TTT GTG GAT TAT TGG ACC AAA AGT TCT TCT GAA ATT TTT 3465
 660 S D T L F V D F V D Y W T K S S E I F 679
 3466 AAA ATG CTC AAG GAA CAT CTC TCT GGA CAC ATT GTT AAG gtataccaattgtgtaattgtaataaca 3532
 680 K M L K E H L S G H I V K 692
 3533 ctaatgaaactag ATA GGA AAT TCT CAA TAC CTT CAA AAA GTT GGT ATC CCT CAG GGC TCA 3593
 693 I G N S Q Y L Q K V G I P Q G S 708
 3594 ATT CTG TCA TCT TTT TTG TGT CAT TTC TAT ATG GAA GAT TTG ATT GAT GAA TAC CTA TCG 3653
 709 I L S S F L C H F Y M E D L I D E Y L S 728
 3654 TTT ACG AAA AAG AAA GGA TCA GTG TTG TTA CGA GTA GTC GAC GAT TTC CTC TTT ATA ACA 3713
 729 F T K K G S V L L R V V D D F L F I T 748
 3714 GTT AAT AAA AAG GAT GCA AAA AAA TTT TTG AAT TTA TCT TTA AGA G gtgagttgctgtcattcc 3777
 749 V N K K D A K K F L N L S L R G 764
 3778 taagtcttaaccgttgaag GA TTT GAG AAA CAC AAT TTT TCT ACG AGC CTG GAG AAA ACA GTA 3840
 765 F E K H N F S T S L E K T V 778
 3841 ATA AAC TTT GAA AAT AGT AAT GGG ATA ATA AAC AAT ACT TTT TTT AAT GAA AGC AAG AAA 3900
 779 I N F E N S N G I I N N T F F N E S K K 798

FIG. 15
(CONTINUED)

3901 AGA ATG CCA TTC TTC GGT TTC TCT GTG AAC ATG AGG TCT CTT GAT ACA TTG TTA GCA TGT 3960
 799 R M P F F G F S V N M R S L D T L L A C 818
 3961 CCT AAA ATT GAT GAA GCC TTA TTT AAC TCT ACA TCT GTA GAG CTG ACG AAA CAT ATG GGG 4020
 819 P K I D E A L F N S T S V E L T K H M G 838
 4021 AAA TCT TTT TAC AAA ATT CTA AG gtatactgtgtaactgaataatagctgacaaataatcag A TCG 4089
 839 K S F F Y K I L R S 848
 4090 AGC CTT GCA TCC TTT GCA CAA GTA TTT ATT GAC ATT ACC CAC AAT TCA AAA TTC AAT TCT 4149
 849 S L A S F A Q V F I D I T H N S K F N S 868
 4150 TGC TGC AAT ATA TAT AGG CTA GGA TAC TCT ATG TGT ATG AGA GCA CAA GCA TAC TTA AAA 4209
 869 C C N I Y R L G Y S M C M R A Q A Y L K 888
 4210 AGG ATG AAG GAT ATA TTT ATT CCC CAA AGA ATG TTC ATA ACG G gtgagtacttattttaactaga 4274
 889 R M K D I F I P Q R M F I T D 903
 4275 aaagtcattaattaaccttag AT CTT TTG AAT GTT ATT GGA AGA AAA ATT TGG AAA AAG TTG GCC 4339
 904 L L N V I G R K I W K K L A 917
 4340 GAA ATA TTA GGA TAT ACG AGT AGG CGT TTC TTG TCC TCT GCA GAA GTC AAA TG gtacgtgtc 4401
 918 E I L G Y T S R R F L S S A E V K W 935
 4402 ggtctcgagacttcagcaaatattgacacatcag G CTT TTT TGT CTT GGA ATG AGA GAT GGT TTG AAA 4468
 936 L F C L G M R D G L K 946
 4469 CCC TCT TTC AAA TAT CAT CCA TGC TTC GAA CAG CTA ATA TAC CAA TTT CAG TCA TTG ACT 4528
 947 P S F K Y H P C F E Q L I Y Q F Q S L T 966
 4529 GAT CTT ATC AAG CCG CTA AGA CCA GTT TTG CGA CAG GTG TTA TTT TTA CAT AGA ATA 4588
 967 D L I K P L R P V L R Q V L F L H R R I 986
 4589 GCT GAT TAA tgtcattttcaatttattatatatacatccttttattactggtgtctttaacaatatattactaagtata 4665
 987 A D * 989

FIG. 15
(CONTINUED)

4666 gctgaccccccaagcagcatactataggatttctagtaaaagtataatctcgttattagttttgattgacttgtct 4745
4746 ttatccttatacttttaagaaagattgacagtgggttgctgactactgcccacatgcccatataaacgggagtggttaaac 4825
4826 ttaaaagtatacatgaggctaatactccttctcaatagaataagaaagtgggttttctataatgaataatgcccgcacta 4905
4906 atgcaaaaagacgaagattatcttctaaacaaggggattaaagcattatccgaaggaaaagagagtaatataccagtggt 4985
4986 gttgaagaaagcaaggataatttggaaacaagcttctgcagatgacaggctaaattttggtgaccgaaattttggtaaaagc 5065
5066 ccagggttatccatgggtggccggccttgctactgagacgaaaaaactaaaggatagtttgaataactaataagctcattta 5145
5146 atgtccttataaagggttttgggttttccctgacttcaattttgcatgggtgaaaagaaatagtggttaagccattattggat 5225
5226 tccgaaatagccaaaatttcttgggttccctcaaacggaagtcataaagaacttattgaagccttatagggttcaaaaactcc 5305
5306 tcctgatttaaggagggaatcttccaccgatagggaatggatagcttatcagctgctgaggagaagccataatttttgc 5385
5386 aaaaaagaaaatatcattgggagacatctcttgatgaatcagatgcggagagtatctccagcggatccttgatgtcaata 5465
5466 acttctatttctgaaatgtatggtcctactgtcgcttcgacttctcgtagctctacgcagtttaagtgaaccacaaagggtacc 5544

FIG. 15
(CONTINUED)

```

1 gcagcgctgc gtctctgctgc gcacgtggga agccctggcc ccggccaccc ccgcgatgcc
61 gcgcgctccc cgctgccgag ccgtgcgctc cctgctgcgc agccactacc gccgagtgct
121 gccgctggcc acgttcgtgc ggcgcctggg gccccagggc tggcggtggg tgcagcgccg
181 ggacccggcg gctttccgcg cgctgggtgg ccagtgccct gtgtgcgtgc cctgggacgc
241 acggccgccc cccgcgcgcc cctccttccg ccaggtgtcc tgcctgaagg agctggtggc
301 ccgagtgctg cagaggctgt gcgagcgcg cgcaagaac gtgctggcct tcggcttcgc
361 gctgctggac gggggccgcg gggggccccc cgaggccttc accaccagcg tgcgcagcta
421 cctgccccaa acggtgaccg acgcactgcg ggggagcggg gcgtgggggg tgctgctgcg
481 ccgctggggc gacgacgtgc tggttcacct gctggcacgc tgcgcgtctt ttgtgctggt
541 ggctcccagc tgcgcctacc aggtgtgcgg gccgcgctg taccagctcg gccgctgccac
601 tcaggcccgg cccccgccac acgctagtgg accccgaagg cgtctgggat gccgaacgggc
661 ctggaaccat agcgtcaggg aggcgggggt cccctggggc ctgccagccc cgggtgagag
721 gagggcgggg ggcagtgcc a gccaagtct gccgttgccc aagaggccca ggcgtggcgc
781 tgccccctgag ccggagcgga cgcccggttg gcagggttcc tgggcccacc cgggagcac
841 gcgtggaccg agtgaccgtg gtttctgtgt ggtgtcacct gccagacccg ccgaagaagc
901 cacctctttg gaggggtgcg tctctggcac gcgccactcc caccatccg tgggcccaca
961 gcaccacgcg gggcccccat ccacatcgcg gccaccagct ccctgggaca cgccttgctc
1021 cccggtgtac gccgagacca agcacttctc ctactcagct ggcgctcgga ggctcgtgga
1081 gcccctcctt ctactcagct ctctgagccc cagcctgact gccagggact cccccgcct
1141 gaccatcttt ctgggttcca ggcctggat gcccctggat gtttctggag ctgcttggga accaccgca
1201 gccccagcgc tactggcaaa tgcggccccct ctgcccgtg cgagctgcgg tcacccagc
1261 gtgccccctac ggggtgctcc tcaagacgca ctgcccgtg gggctctgtg agggaggaga ggcaggtgta
1321 agccagcttc tgtgcccggt tgcagctgct ccgccagcac ggcctctggg gctccaggca
1381 cacagacccc cgctgcctgg tgcgcccgtt ggtgccccca tccctgggga agcatgccaa
1441 cggcttcgtg cgggcctgcc tgcgcccgtt ggtgccccca tccctgggga agcatgccaa
1501 caacgaacgc cgcttctcct cgtggaagat gacgctgcgg gactgcgctt gactgcgctt
1561 gctctcgtcg caggagctga ttccggccgc agagcacctg ctgctgaggg agatcctggc
1621 gagcccaggg gttggctgtg tgagtgtgta cgtcgtcgag ctgctcaggt ctttctttta
1681 caagtctctg cactggctga tgaagtgtgta gctcttttcc taccggaaga gtgtctggag
1741 tgtcacggag accacgtttc tcaagacgca ctgaaagagg gtgcagctgc ccagactccg
1801 caagtgtcaa agcattggaa atcggaagc caggcccgcg atggactacg tgcgtgggagc
1861 ggaagcagag gtcaggcagc ggctgcggcc gattgtgaac tcgaggggtg aggcactgtt
1921 cttcatcccc aagcctgacg agagggccga ggcgtctacc ccccgcctc ctggtgggg
1981 cagaacgttc cgagagaaa gggcgccgac cttcgtgctg cgtgtgcggg cccaggaccc
2041 cagcgtgctc aactacgag cctggcgcac tcaagtgga tgtgacgggc gcgtacgaca
2101 cctggacgat atccacaggg tgcggcagat catcaaacc cagaacacgt actgctgcg
2161 gccgcctgag ctgtactttg tgcggcagat tgggacgct cgaaggcct tcaagaccca
2221 ggacaggctc acggaggtca aggccgccc tgggacgct ttcgtggctc acctgagga
2281 tcggtatgcc gtggtccaga tccagcgtga cagcagagc tccctcctga atgaggccag
2341 cgtctctacc ttgacagacc ccgtcgtcat catgtgccac cagcccgctg gcacagggg
2401 gaccagcccc ctgagggatg tccacgtctt aggggatccc gcagggtccc atcctctcca cctgctctg
2461 cagtggcctc ttcgacgtct tggagaacaa gctgtttgct ggtgacacct cacctcacc
2521 caagtccctc gtccagtgc ttttcttgtt tgagtatggc tgcgtggtga acttgcgga
2581 gactcgacac tacggcgaca tagaagacga ggccttgggt ggcacggctt ttgttcagat
2641 gctcctgcgt ttggtggatg cctggctgcg cctgctgctg gatacccga cctggaggt
2701 ctctctcagg accctggctc atgcccggac ctccatcaga gccagtctca ccttcaaccg
2761 gacagtgggt aacttccctg acatgcgtcg caaactcttt ggggtcttgc ggctgaagt
2821 gccggcccac ggcctattcc tgcaggtgaa cagcctccag acggtgtgca ccaacatcta
2881 gcagagcgac tactccagct cgtacaggtt tcacgcatgt gtgctgcagc tcccatttca
2941 cggcttcaag gctgggagga ccacattttt cctgcgctc atctctgaca cggcctccct
3001 tcacagcctg tttctggatt ccaagaacgc agggatgtcg ctggggggcca agggcgccgc
3061 caagatcctc ctgctgcagg ccgtgcagtg gctgtgccac caagcatcc tgctcaagct
3121 tcagcaagtt tggagaacc cgtgtcacct cctgggggtc ctcaggacag cccagacgca
3181 ctgctactcc atcctgaaag ggacgacgct gactgccctg gagggcgag ccaaccgggc
3241 cggccctctg cctccgagg gacgacgct ccaccttggg ctgatggcca cccgccaca
3301 gactcgacac agtgtcacct gtcacgcccg gctctacgt ccagggaggg agggggcgcc
3361 gctgagtcgg aagctcccg gaggagctgt tgggagctg aggcctgag cagagcctg
3421 actgccctca gacttcaaga tgggagctgt gggcctgag tggcgtcggt cccacacagg
3481 gagcagacac cagcagccct tgggagctgt aggcctgag tggcgtcggt cccacacagg
3541 cacacccagg cccgcaccgc gtcttcaact cccacacagg tggcgtcggt cccacacagg
3601 catgtccggc tgaaggctga gtagcccggt tggcgtcggt cccacacagg tggcgtcggt
3661 gagtgctccg cacacctgcc gtagcccggt tggcgtcggt cccacacagg tggcgtcggt
3721 gggccagctt ttcctacca cctgcgcctt ggcctccttt cccacacagg tggcgtcggt
3781 ccagattcgc cattgttcac cctgcgcctt ggcctccttt cccacacagg tggcgtcggt
3841 aggtggagac cctgagaagg cctgcacctg ccttgggaat tggagtgac caaaggtgtg
3901 cctgtacac aggcgaggac cctgcacctg gatgggggtc cctgtgggtc aaattggggg
3961 gaggtgctgt gggagtaaaa tactgaatat atgagttttt cagttttgaa aaaaa

```

FIG. 16

MPRAPRCRAVRSLLRSHYREVLPLATFVRRLGPQGWRLVQRGDP
 AAFRALVAQCLVCVPWDARPPPAAPSFQVSCLELVARVLQRL
 CERGAQNVLAFGFALLDGARGGPPEAFTTSVRSYLPNTVTDALR
 GSGAWGLLLRRVGDVLLVHLLARCALFVLVAPSCAYQVCGPPLY
 QLGAATQARPPPHASGPRRRLGCERAWNHSVREAGVPLGLPAPG
 ARRCGSASRSLPLPKRPRRGAPEPERTPVQGSWAHPGRTRG
 PSDRGFCVVSAPPAEEATSLEGALSGTRHSHPSVGRQHHAGPP
 STSRPPRPWDTPCPPVYAETKHFYSSGDKEQLRPSFLLSSLRP
 SLTGARRLVETIFLGSRPWMPGTFRRLPRLPQRYWQMRPLFLEL
 LGNHAQCPYGVLLKTHCPLRAAVTPAAGVCAREKPGGSVAAPPEE
 EDTDPRRLVQLLRQHSSPWQVYGFVRACLRLVPPGLWGSRHNE
 RRFLRNTKKFISLGKHAKLSLQELTWKMSVRDCAWLRRSPGVGC
 VPAAEHLREEILAKFLHWLMSVYVVELLSFFVYTETTFQKNR
 LFFYRKS VWSKLQSIGIRQHLKRVQLRELSEAEVRQHREARPA
 LTSRLRFIPKPDGLRPVNM DYVVGARTFRREKRAERLTSRVKA
 LFSVLNYERARRPGLLGASVLGLDDIHRWRTFVLRVRAQDPPP
 ELYFVKVDVTGAYDTIPQDRLTEVIASIIKPNITYCVRRYAVVQ
 KAAHGHRKAFKSHVSTLTDLPYMRQFVAHLQETSPLRDAVVI
 EQSSSLNEASSGLFDVFLRFMCHHAVRIRGKSYVQCQGI PQGSI
 LSTLLCSLCYGD MENKLFAGIRRDGLLLRLVDDFLLVTPHLTHA
 KTFRLTLVRGVPEYGC VVNLRKTVVNFVVEDEALGGTAFVQMPA
 HGLFPWCGLLLDTRTLEVQSDYSSYARTSIRASLTFNRGFKAGR
 NMRRKLFGLRLKCHSLF LDLQVNSLQTVCTNIYKILLQAYRF
 HACVLQLPFHQVWKNPTFFLRVISDTASLCYSILKAKNAGMSL
 GAKGAAGPLPSEAVQWLCHQAFLLKLTRHRVTYVPLLGLSLRTAQ
 TQLSRKLPGTTLTALEAAANPALPSDFKTILD

FIG. 17

GGCCAAGTTCCTGCACTGGCTGATGAGTGTGTACGTCGTCGAGCTGCTCAGGTCTTTCTT
 TTATGTCACGGAGACCACGTTTCAAAAAGAACAGGCTCTTTTCTACCGGAAGAGTGTCTG
 GAGCAAGTTGCAAAGCATTTGAATCAGACAGCACTTGAAGAGGTGCAGCTGCGGGAGCT
 GTCGGAAGCAGAGGTGAGGCAGCATCGGGAAGCCAGGCCCGCCCTGCTGACGTCCAGACT
 CCGCTTTCATCCCCAAGCCTGACGGGCTGCGGCCGATTGTGAACATGGACTACGTCGTGGG
 AGCCAGAAGCTTCCGAGAGAAAAGAGGGCCGAGCGTCTCACCTCGAGGGTGAAGGCACT
 GTTCAGCGTGTCTCAACTACGAGCGGGCGCGGCCCGCCCTCTGGGCGCCTCTGTGTCT
 GGGCCTGGACGATATCCACAGGGCCTGGCGCACCTTCGTGTGCGTGTGCGGGCCAGGA
 CCGCGCCGCTGAGCTGTACTTTGTCAAGGTGGATGTGACGGGCGCGTACGACACCATCCC
 CCAGGACAGGCTCACGGAGGTATCGCCAGCATCATCAAACCCAGAACACGTACTGCGT
 GCGTCGGTATGCCGTGGTCCAGAAGGCCCGCCATGGGCACGTCCGCAAGGCCTTCAAGAG
 CCACGTCTTACGTCCAGTGCCAGGGGATCCCGCAGGGCTCCATCTCTCCACGCTGTCT
 GCAGCCTGTGTCTACGGCGACATGGAGAACAAGCTGTTTGCGGGGATTGCGCGGGACGGG
 TGCTCTCTGCGTTTGGTGGATGATTTCTGTGTGGTGACACCTCACCTCACCCGCGAAAA
 CCTTCTCTCAGGACCTGCTCCGAGGTGTCCCTGAGTATGGCTGCGTGGTGAACCTGCGGA
 AGACAGTGGTGAACCTCCCTGTAGAAGACGAGGCCCTGGGTGGCACGGCTTTTGTTCAGA
 TGCCGGCCCCACGGCCTATTCCCTTGGTGCGGCTGTGTGATACCCGGACCTTGGAGG
 TGCAGAGCGACTACTCCAGCTATGCCCGGACCTCCATCAGAGCCAGTCTCACCTTCAACC
 GCGGCTTCAAGGCTGGGAGGAACATGCGTCGCAAACTCTTTGGGGTCTTTCGGGCTGAAGT
 GTCACAGCCTGTTTCTGGATTGTCAGGTGAACAGCCTCCAGACGGTGTGCACCAACATCT
 ACAAGATCTCTCTGCTGCAGGCGTACAGGTTTCACGCATGTGTGCTGCAGCTCCCATTTT
 ATCAGCAAGTTTGAAGAACCACATTTTCTCTGCGCGTCATCTCTGACACGGCCTCC
 TCTGCTACTCCATCTTGAAGCCAAGAAGCAGGGATGTGCTGGGGGCCAAGGGCGCG
 CCGGCC7TCTGCCCTCCGAGGCGGTGACGTGGCTGTGCCACCAAGCATTCCTGCTCAAGC
 TGACTCGACACCGGTGTACCTACGTGCCACTCTCTGGGGTCACTCAGGACAGCCAGACGC
 AGCTGAGTCGGAAGCTCCCGGGGACGACGTGACTGCCCTGGAGGCCGAGCCAACCCGG
 CACTGCCCTCAGACTTCAAGACCATCTTGGACTGATGGCCACCCGCCACAGCCAGGCCG
 AGAGCAGACACCAGCAGCCCTGTACGCCGGGCTCTACGTCCCAGGGAGGGGAGGGCGG
 CCACACCCAGGCCTGCACCGCTGGGAGTCTGAGGCTGAGTGAGTGTGGCCGAGGCCT
 GCAATGTCGGCTGAAGGCTGAGTGTCCGGCTGAGGCTGAGCGAGTGTCCAGCCAAGGGC
 TGAGTGTCCAGCACACCTGCCGTCTTCACTTCCCAAGGCTGGCGCTCGGCTCCACCCC
 AGGGCAGCTTTTCTTCAACAGGAGCCCGGCTTCACTCCCAACATAGGAATAGTCCATC
 CCCAGATTGCCCATGTTTACCCCTCGCCCTGCCCTCTTTCCTTCCACCCACCATC
 CAGGTGGAGACCCCTGAGAAGGACCTGGGAGCTCTGGGAATTTGGAGTGACCAAAGGTGT
 GCCCTGTACACAGGCGAGGACCTGCACCTGGATGGGGTCCCTGTGGGTCAAATTGGGG
 GGAGGTGCTGTGGGAGTAAATACTGAATATATGAGTTTTCAGTTTGTG0AAAAA
 AAAAAAAAAAAAAA

FIG. 18

MetSerValTyrValValGluLeuLeuArgSerPhePhe
TyrValThrGluThrThrPheGlnLysAsnArgLeuPhe
PheTyrArgLysSerValTrpSerLysLeuGlnSerIle
GlyIleArgGlnHisLeuLysArgValGlnLeuArgGlu
LeuSerGluAlaGluValArgGlnHisArgGluAlaArg
ProAlaLeuLeuThrSerArgLeuArgPheIleProLys
ProAspGlyLeuArgProIleValAsnMetAspTyrVal
ValGlyAlaArgThrPheArgArgGluLysArgAlaGlu
ArgLeuThrSerArgValLysAlaLeuPheSerValLeu
AsnTyrGluArgAlaArgArgProGlyLeuLeuGlyAla
SerValLeuGlyLeuAspAspIleHisArgAlaTrpArg
ThrPheValLeuArgValArgAlaGlnAspProProPro
GluLeuTyrPheValLysValAspValThrGlyAlaTyr
AspThrIleProGlnAspArgLeuThrGluValIleAla
SerIleIleLysProGlnAsnThrTyrCysValArgArg
TyrAlaValValGlnLysAlaAlaHisGlyHisValArg
LysAlaPheLysSerHisValLeuArgProValProGly
AspProAlaGlyLeuHisProLeuHisAlaAlaLeuGln
ProValLeuArgArgHisGlyGluGlnAlaValCysGly
AspSerAlaGlyArgAlaAlaProAlaPheGlyGly

														1
														met
GCAGCGCTGCGTCCTGCTGCGCACGTGGGAAGCCCTGGCCCCGCCACCCCGCG														ATG
														10
pro	arg	ala	pro	arg	cys	arg	ala	val	arg	ser	leu	leu	arg	ser
CCG	CGC	GCT	CCC	CGC	TGC	CGA	GCC	GTG	CGC	TCC	CTG	CTG	CGC	AGC
														20
his	tyr	arg	glu	val	leu	pro	leu	ala	thr	phe	val	arg	arg	leu
CAC	TAC	CGC	GAG	GTG	CTG	CCG	CTG	GCC	ACG	TTC	GTG	CGG	CGC	CTG
														30
														40
gly	pro	gln	gly	trp	arg	leu	val	gln	arg	gly	asp	pro	ala	ala
GGG	CCC	CAG	GGC	TGG	CGG	CTG	GTG	CAG	CGC	GGG	GAC	CCG	GCG	GCT
														50
phe	arg	ala	leu	val	ala	gln	cys	leu	val	cys	val	pro	trp	asp
TTC	CGC	GCG	CTG	GTG	GCC	CAG	TGC	CTG	GTG	TGC	GTG	CCC	TGG	GAC
														60
														70
ala	arg	pro	pro	pro	ala	ala	pro	ser	phe	arg	gln	val	ser	cys
GCA	CGG	CCG	CCC	CCC	GCC	GCC	CCC	TCC	TTC	CGC	CAG	GTG	TCC	TGC
														80
leu	lys	glu	leu	val	ala	arg	val	leu	gln	arg	leu	cys	glu	arg
CTG	AAG	GAG	CTG	GTG	GCC	CGA	GTG	CTG	CAG	AGG	CTG	TGC	GAG	CGC
														90
														100
gly	ala	lys	asn	val	leu	ala	phe	gly	phe	ala	leu	leu	asp	gly
GGC	GCG	AAG	AAC	GTG	CTG	GCC	TTC	GGC	TTC	GCG	CTG	CTG	GAC	GGG
														110
ala	arg	gly	gly	pro	pro	glu	ala	phe	thr	thr	ser	val	arg	ser
GCC	CGC	GGG	GGC	CCC	CCC	GAG	GCC	TTC	ACC	ACC	AGC	GTG	CGC	AGC

FIG. 20

130
 tyr leu pro asn thr val thr asp ala leu arg gly ser gly ala
 TAC CTG CCC AAC ACG GTG ACC GAC GCA CTG CGG GGG AGC GGG GCG

140 150
 trp gly leu leu leu arg arg val gly asp asp val leu val his
 TGG GGG CTG CTG CTG CGC CGC GTG GGC GAC GAC GTG CTG GTT CAC

160
 leu leu ala arg cys ala leu phe val leu val ala pro ser cys
 CTG CTG GCA CGC TGC GCG CTC TTT GTG CTG GTG GCT CCC AGC TGC

170 180
 ala tyr gln val cys gly pro pro leu tyr gln leu gly ala ala
 GCC TAC CAG GTG TGC GGG CCG CCG CTG TAC CAG CTC GGC GCT GCC

190
 thr gln ala arg pro pro pro his ala ser gly pro arg arg arg
 ACT CAG GCC CGG CCC CCG CCA CAC GCT AGT GGA CCC CGA AGG CGT

200 210
 leu gly cys glu arg ala trp asn his ser val arg glu ala gly
 CTG GGA TGC GAA CGG GCC TGG AAC CAT AGC GTC AGG GAG GCC GGG

220
 val pro leu gly leu pro ala pro gly ala arg arg arg gly gly
 GTC CCC CTG GGC CTG CCA GCC CCG GGT GCG AGG AGG CGC GGG GGC

230 240
 ser ala ser arg ser leu pro leu pro lys arg pro arg arg gly
 AGT GCC AGC CGA AGT CTG CCG TTG CCC AAG AGG CCC AGG CGT GGC

250
 ala ala pro glu pro glu arg thr pro val gly gln gly ser trp
 GCT GCC CCT GAG CCG GAG CGG ACG CCC GTT GGG CAG GGG TCC TGG

260 270
 ala his pro gly arg thr arg gly pro ser asp arg gly phe cys
 GCC CAC CCG GGC AGG ACG CGT GGA CCG AGT GAC CGT GGT TTC TGT

280
 val val ser pro ala arg pro ala glu glu ala thr ser leu glu
 GTG GTG TCA CCT GCC AGA CCC GCC GAA GAA GCC ACC TCT TTG GAG

290 300
 gly ala leu ser gly thr arg his ser his pro ser val gly arg
 GGT GCG CTC TCT GGC ACG CGC CAC TCC CAC CCA TCC GTG GGC CGC

310
 gln his his ala gly pro pro ser thr ser arg pro pro arg pro
 CAG CAC CAC GCG GGC CCC CCA TCC ACA TCG CGG CCA CCA CGT CCC

320 330
 trp asp thr pro cys pro pro val tyr ala glu thr lys his phe
 TGG GAC ACG CCT TGT CCC CCG GTG TAC GCC GAG ACC AAG CAC TTC

FIG. 20
(CONTINUED)

25/103

```

                                340
leu tyr ser ser gly asp lys glu gln leu arg pro ser phe leu
CTC TAC TCC TCA GGC GAC AAG GAG CAG CTG CGG CCC TCC TTC CTA

                                350
leu ser ser leu arg pro ser leu thr gly ala arg arg leu val
CTC AGC TCT CTG AGG CCC AGC CTG ACT GGC GCT CGG AGG CTC GTG

                                360
                                370
glu thr ile phe leu gly ser arg pro trp met pro gly thr pro
GAG ACC ATC TTT CTG GGT TCC AGG CCC TGG ATG CCA GGG ACT CCC

                                380
                                390
arg arg leu pro arg leu pro gln arg tyr trp gln met arg pro
CGC AGG TTG CCC CGC CTG CCC CAG CGC TAC TGG CAA ATG CGG CCC

                                400
leu phe leu glu leu leu gly asn his ala gln cys pro tyr gly
CTG TTT CTG GAG CTG CTT GGG AAC CAC GCG CAG TGC CCC TAC GGG

                                410
                                420
val leu leu lys thr his cys pro leu arg ala ala val thr pro
GTG CTC CTC AAG ACG CAC TGC CCG CTG CGA GCT GCG GTC ACC CCA

                                430
ala ala gly val cys ala arg glu lys pro gln gly ser val ala
GCA GCC GGT GTC TGT GCC CGG GAG AAG CCC CAG GGC TCT GTG GCG

                                440
                                450
ala pro glu glu glu asp thr asp pro arg arg leu val gln leu
GCC CCC GAG GAG GAG GAC ACA GAC CCC CGT CGC CTG GTG CAG CTG

                                460
leu arg gln his ser ser pro trp gln val tyr gly phe val arg
CTC CGC CAG CAC AGC AGC CCC TGG CAG GTG TAC GGC TTC GTG CGG

                                470
                                480
ala cys leu arg arg leu val pro pro gly leu trp gly ser arg
GCC TGC CTG CGC CGG CTG GTG CCC CCA GGC CTC TGG GGC TCC AGG

                                490
his asn glu arg arg phe leu arg asn thr lys lys phe ile ser
CAC AAC GAA CGC CGC TTC CTC AGG AAC ACC AAG AAG TTC ATC TCC

                                500
                                510
leu gly lys his ala lys leu ser leu gln glu leu thr trp lys
CTG GGG AAG CAT GCC AAG CTC TCG CTG CAG GAG CTG ACG TGG AAG

                                520
met ser val arg asp cys ala trp leu arg arg ser pro gly val
ATG AGC GTG CGG GAC TGC GCT TGG CTG CGC AGG AGC CCA GGG GTT

                                530
                                540
gly cys val pro ala ala glu his arg leu arg glu glu ile leu
GGC TGT GTT CCG GCC GCA GAG CAC CGT CTG CGT GAG GAG ATC CTG

```

FIG. 20
(CONTINUED)

550
 ala lys phe leu his trp leu met ser val tyr val val glu leu
 GCC AAG TTC CTG CAC TGG CTG ATG AGT GTG TAC GTC GTC GAG CTG

560 570
 leu arg ser phe phe tyr val thr glu thr thr phe gln lys asn
 CTC AGG TCT TTC TTT TAT GTC ACG GAG ACC ACG TTT CAA AAG AAC

580
 arg leu phe phe tyr arg lys ser val trp ser lys leu gln ser
 AGG CTC TTT TTC TAC CGG AAG AGT GTC TGG AGC AAG TTG CAA AGC

590 600
 ile gly ile arg gln his leu lys arg val gln leu arg glu leu
 ATT GGA ATC AGA CAG CAC TTG AAG AGG GTG CAG CTG CGG GAG CTG

610
 ser glu ala glu val arg gln his arg glu ala arg pro ala leu
 TCG GAA GCA GAG GTC AGG CAG CAT CGG GAA GCC AGG CCC GCC CTG

620 630
 leu thr ser arg leu arg phe ile pro lys pro asp gly leu arg
 CTG ACG TCC AGA CTC CGC TTC ATC CCC AAG CCT GAC GGG CTG CGG

640
 pro ile val asn met asp tyr val val gly ala arg thr phe arg
 CCG ATT GTG AAC ATG GAC TAC GTC GTG GGA GCC AGA ACG TTC CGC

650 660
 arg glu lys arg ala glu arg leu thr ser arg val lys ala leu
 AGA GAA AAG AGG GCC GAG CGT CTC ACC TCG AGG GTG AAG GCA CTG

670
 phe ser val leu asn tyr glu arg ala arg arg pro gly leu leu
 TTC AGC GTG CTC AAC TAC GAG CGG GCG CGG CGC CCC GGC CTC CTG

680 690
 gly ala ser val leu gly leu asp asp ile his arg ala trp arg
 GGC GCC TCT GTG CTG GGC CTG GAC GAT ATC CAC AGG GCC TGG CGC

700
 thr phe val leu arg val arg ala gln asp pro pro pro glu leu
 ACC TTC GTG CTG CGT GTG CGG GCC CAG GAC CCG CCG CCT GAG CTG

710 720
 tyr phe val lys val asp val thr gly ala tyr asp thr ile pro
 TAC TTT GTC AAG GTG GAT GTG ACG GGC GCG TAC GAC ACC ATC CCC

730
 gln asp arg leu thr glu val ile ala ser ile ile lys pro gln
 CAG GAC AGG CTC ACG GAG GTC ATC GCC AGC ATC ATC AAA CCC CAG

740 750
 asn thr tyr cys val arg arg tyr ala val val gln lys ala ala
 AAC ACG TAC TGC GTG CGT CGG TAT GCC GTG GTC CAG AAG GCC GCC

FIG. 20
(CONTINUED)

27/103

760
his gly his val arg lys ala phe lys ser his val leu arg pro
CAT GGG CAC GTC CGC AAG GCC TTC AAG AGC CAC GTC CTA CGT CCA

770 780
val pro gly asp pro ala gly leu his pro leu his ala ala leu
GTG CCA GGG GAT CCC GCA GGG CTC CAT CCT CTC CAC GCT GCT CTG

790
gln pro val leu arg arg his gly glu gln ala val cys gly asp
CAG CCT GTG CTA CGG CGA CAT GGA GAA CAA GCT GTT TGC GGG GAT

800 807
ser ala gly arg ala ala pro ala phe gly gly OP
TCG GCG GGA CGG GCT GCT CCT GCG TTT GGT GGA TGA TTTCTTGTGGT

GACACCTCACCTCACCCACGCGAAAACCTTCCTCAGGACCCTGGTCCGAGGTGTCCCTGA
GTATGGCTGCGTGGTGAACCTTGCGGAAGACAGTGGTGAACCTTCCCTGTAGAAGACGAGGC
CCTGGGTGGCACGGCTTTTGTTCAGATGCCGGCCACGGCCTATTCCCCTGGTGGCGCCT
GCTGCTGGATAACCCGACCCCTGGAGGTGCAGAGCGACTACTCCAGCTATGCCCGGACCTC
CATCAGAGCCAGTCTCACCTTCAACCGCGGCTTCAAGGCTGGGAGGAACATGCGTCGCAA
ACTCTTTGGGGTCTTGCGGCTGAAGTGTACAGCCTGTTTCTGGATTGTGAGGTGAACAG
CCTCCAGACGGTGTGCACCAACATCTACAAGATCCTCCTGCTGCAGGCGTACAGGTTTCA
CGCATGTGTGCTGCAGCTCCCATTTCATCAGCAAGTTTGAAGAACCCACATTTTTCCT
GCGCGTCATCTCTGACACGGCCTCCCTCTGCTACTCCATCCTGAAAGCCAAGAACGCAGG
GATGTCGCTGGGGGCCAAGGGCGCCGCCGCCCTCTGCCCTCCGAGGCCGTGCAGTGGCT
GTGCCACCAAGCATTTCCTGCTCAAGCTGACTCGACACCGTGTACCTACGTGCCACTCCT
GGGGTCACTCAGGACAGCCCAGACGCAGCTGAGTCGGAAGCTCCCGGGGACGACGCTGAC
TGCCCTGGAGGCCGAGCCAACCCGGCACTGCCCTCAGACTTCAAGACCATCCTGGACTG
ATGGCCACCCGCCCACAGCCAGGCCGAGAGCAGACACCAGCAGCCCTGTCACGCCGGGCT
CTACGTCCCAGGGAGGGAGGGGCGGCCACACCCAGGCCCGCACCGCTGGGAGTCTGAGG
CCTGAGTGAGTGTGTTGGCCGAGGCCTGCATGTCCGGCTGAAGGCTGAGTGTCCGGCTGAG
GCCTGAGCGAGTGTCCAGCCAAGGGCTGAGTGTCCAGCACACCTGCCGTCTTCACTTCCC
CACAGGCTGGCGCTCGGCTCCACCCAGGGCCAGCTTTTCTCACCAGGAGCCCGGCTTC
CACTCCCCACATAGGAATAGTCCATCCCCAGATTGCGCCATTGTTACCCCTCGCCCTGCC
CTCCTTTGCCTTCCACCCCCACCATCCAGGTGGAGACCCTGAGAAGGACCCTGGGAGCTC
TGGAATTTGGAGTGACCAAGGTGTGCCCTGTACACAGGCGAGGACCCTGCACCTGGAT
GGGGTCCCTGTGGGTCAAATTGGGGGGAGGTGCTGTGGGAGTAAAATACTGAATATATG
AGTTTTTCAGTTTTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

FIG. 20
(CONTINUED)

```

1  CCATGGGACCCACTGCAGGGGCAGCTGGGAGGCTGCAGGCTTCAGGTCCCAGTGGGGTTG
   GGTACCCTGGGTGACGTCCCCGTGACCCCTCCGACGTCCGAAGTCCAGGGTCACCCCAAC

61  CCATCTGCCAGTAGAAACCTGATGTAGAATCAGGGCGCGAGTGTGGACACTGTCC'TGAAT
   GGTAGACGGTCATCTTTGGACTACATCTTAGTCCCGCGCTCACACCTGTGACAGGACTTA

121  CTCAATGTCTCAGTGTGTGCTGAAACATGTAGAAATTAAAGTCCATCCCTCCTACTCTAC
   GAGTTACAGAGTCACACACGACTTTGTACATCTTTAATTTTCAGGTAGGGAGGATGAGATG

181  TGGGATTGAGCCCCCTTCCCTATCCCCCCCCAGGGGCAGAGGAGTTCTCTCACTCCTGTG
   ACCCTAACTCGGGGAAGGGATAGGGGGGGGTCCCCGTCTCCTCAAGGAGAGTGAGGACAC

241  GAGGAAGGAATGATACTTTTGTTATTTTTCAGTGTGCTGGTACTGAATCCACTGTTTCATTTG
   CTCTTCCTTACTATGAACAATAAAAAGTGACGACCATGACTTAGGTGACAAAGTAAAC

                                     *****
301  TTGGTTTGTGTTTGTGTTTGTGTTTGGAGAGCGGTTTCACTCTTGTGCTCAGGCTGGAGGGAG
   AACCAACAAACAAAACAAAACCTCTCCGCCAAAGTGAGAACACGAGTCCGACCTCCCTC

                                     *****
361  TGCAATGGCGCGATCTTGGCTTACTGCAGCCTCTGCCTCCCAGGTTCAAGTGATTCTCCT
   ACGTTACCGCGCTAGAACCGAATGACGTCGGAGACGGAGGGTCCAAGTTCACCTAAGAGGA

                                     alu
                                     *****
421  GCTTCCGCCTCCCATTTGGCTGGGATTACAGGCACCCGCCACCATGCCAGCTAAATTTT
   CGAAGGCGGAGGGTAAACCGACCCTAATGTCCGTGGGCGGTGGTACGGGTGCGATTAAAAA

                                     ==

                                     *****
481  TGTATTTT'TAGTAGAGACGGGGGTGGGGGTGGGGTTCACCATGTTGGCCAGGCTGGTCTC
   ACATAAAAATCATCTCTGCCCCACCCCCACCCCAAGTGGTACAACCGGTCCGACCAGAG

CAP
=====>

                                     *****
541  GAACTTCTGACCTCAGATGATCCACCTGCCTCTGCCTCCTAAAGTGCTGGGATTACAGGT
   CTTGAAGACTGGAGTCTACTAGGTGGACGGAGACGGAGGATTTACGACCCCTAATGTCCA

                                     *****
601  GTGAGCCACCATGCCAGCTCAGAATTTACTCTGTTTAGAAACATCTGGGTCTGAGGTAG
   CACTCGGTGGTACGGGTGAGTCTTAAATGAGACAAATCTTTGTAGACCCAGACTCCATC

                                     CCAAT
                                     *****>
661  GAAGCTCACCCCACTCAAGTGTTGTGGTGT'TTAAGCCAATGATAGAAATTTT'TATGT
   CTTGAGTGGGGTGAGTTCACAACACCACAAAATTCGGTTACTATCTTAAAAAATAACA

721  TGTTAGAACACTCTTGATGTTTTACTGTGATGACTAAGACATCATCAGCTTTTCAAAG
   ACAATCTTGTGAGAACTACAAAATGTGACACTACTGATTCTGTAGTAGTCGAAAAGTTTC

```

FIG. 21

CAP

*****>

781 ACACACTAACTGCACCCATAATACTGGGGTGTCTTCTGGGTATCAGCGATCTTCATTGAA
TGTGTGATTGACGTGGGTATTATGACCCACAGAAGACCCATAGTCGCTAGAAGTAACCT

CAP

841 TGCCGGGAGGCGTTTCTCGCCATGCACATGGTGTAAATTACTCCAGCATAATCTTCTGC
ACGGCCCTCCGCAAAGGAGCGGTACGTGTACCACAATTAATGAGGTCTGATTAGAAGACG

***>

901 TTCCATTTCTTCTCTTCCCTCTTTTAAAATTGTGTTTTCTATGTTGGCTTCTCTGCAGAG
AAGGTAAAGAAGAGAAGGGAGAAAATTTTAAACACAAAAGATACAACCGAAGAGACGTCTC

CAP

*****>

961 AACCAGTGTAAGCTACAACCTTAACCTTTTGTGGAACAAATTTTCAAACCGCCCTTTTGC
TTGGTTCACATTTCGATGTTGAATTGAAAACAACCTTGTTTAAAAGGTTTGGCGGGGAAACG

1021 CCTAGTGGCAGAGACAATTCACAAACACAGCCCTTTAAAAGGCTTAGGGATCACTAAGG
GGATCACCGTCTCTGTTAAGTGTGTGTCGGGAAATTTTCCGAATCCCTAGTGATTCC

1081 GGATTTCTAGAAGAGCGACCCGTAATCCTTAAGTATTTACAAGACGAGGCTAACCTCCAG
CCTAAAGATCTTCTCGCTGGGCATTAGGAATTCATAAATGTTCTGCTCCGATTGGAGGTC

1141 CGAGCGTGACAGCCCAGGGAGGGTGCGAGGCCTGTTCAAATGCTAAGCTTCCATAAATAA
GCTCGCACTGTCTGGGTCCCTCCCACGCTCCGGACAAGTTTACGATTCTGAAGGTATTTATT

1201 AGCAAATTTCTCCGGCAGTTTCTGGAAAGTAGGAAAGGTTAACATTTAAGGTTGCGTTT
TCGTTTAAAGGAGGCCGTCAAAGACCTTTTCATCCTTTCCAATTGTAAATTTCAAACGCAAA

1261 GTTAGCATTTTCAGTGTTTGCCGACCTCAGCTAACAGCATCCCTGCAAGGCCTCGGGAGAC
CAATCGTAAAGTCACAAACGGCTGGAGTCGATTGTCTGCTAGGGACGTTCCGGAGCCCTCTG

1321 CCAGAAGTTTCTCGCCCTTAGATCCAACTTGAGCAACCCGGAGTCTGGATTCTCTGGGA
GGTCTTCAAAGAGCGGGGAATCTAGGTTTGAACTCGTTGGGCCTCAGACCTAAGGACCCT

TopoII

*****>

1381 AGTCCTCAGCTGTCTGCGGTTGTGCCGGGGCCCCAGGTCTGGAGGGGACCAGTGGCCGT
TCAGGAGTCGACAGGACGCCAACACGGCCCCGGGTCCAGACCTCCCCTGGTCACCGGCA

1441 GTGGCTTCTACTGCTGGGCTGGAAGTCGGGCCTCCTAGCTCTGCAGTCCGAGGCTTGGAG
CACCGAAGATGACGACCCGACCTTCAGCCCGGAGGATCGAGACGTCAGGCTCCGAACCTC

1501 CCAGGTGCCTGGACCCCGAGGCTGCCCTCCACCCTGTGCGGGCGGGATGTGACCAGATGT
GGTCCACGGACCTGGGGCTCCGACGGGAGGTGGGACACGCCCGCCCTACACTGGTCTACA

1561 TGGCCTCATCTGCCAGACAGAGTGCCGGGGCCAGGGTCAAGGCCGTTGTGGCTGGTGTG
ACCGGAGTAGACGGTCTGTCTCAGGCCCGGGTCCAGTTCCGGCAACACCGACCACAC

1621 AGGCGCCCGGTGCGCGGCCAGCAGGAGCGCCTGGCTCCATTTCCACCCCTTTCTCGACGG
TCCGCGGGCCACGCGCCGTCGTCTCGCGGACCGAGGTAAAGGGTGGGAAAGAGCTGCC

FIG. 21
(CONTINUED)

1681 GACCGCCCCGGTGGGTGATTAACAGATATTGGGGTGGTTTGCTCATGGTGGGGACCCCTT
 CTGGCGGGGCCACCCACTAATTGTCTATAACCCACCAAACGAGTACCACCCCTGGGGAA
 1741 CGCCGCCTGAGAACCTGCAAAGAGAAATGACGGGCCTGTGTCAAGGAGCCCAAGTCGCGG
 GCGGCGGACTCTTGACGTTTCTCTTTACTGCCCCGACACAGTTCCCTCGGGTTCAGCGCC
 1801 GGAAGTGTGTCAGGGAGGCACTCCGGGAGGTCCCGCGTGCCCGTCCAGGGAGCAATGCGT
 CCTTCACAACGTCCCTCCGTGAGGCCCTCCAGGGCGCACGGGCAGGTCCCTCGTTACGCA
 1861 CCTCGGGTTCGTCCCCAGCCGCGTCTACGCGCCTCCGTCCCTCCCTTCACGTCCGGCATT
 GGAGCCCAAGCAGGGGTGGGCGCAGATGCGCGGAGGCAGGAGGGGAAGTGCAGGCCGTAA
 1921 CGTGGTGCCCCGAGCCCGACGCCCCGCGTJCGGACCTGGAGGCAGCCCTGGGTCTCCGGA
 GCACCACGGGCCTCGGGCTGCGGGGCGCAGGCCTGGACCTCCGTCTGGGACCCAGAGGCCT
 1981 TCAGGCCAGCGGCCAAAGGGTCGCCGCACGCACCTGTTCCAGGGCCTCCACATCATGGC
 AGTCCGGTTCGCCGTTTCCAGCGGCGTGGTGGACAAGGGTCCCGGAGGTGTAGTACCG
 2041 CCCTCCCTCGGGTTACCCACAGCCTAGGCCGATTTCGACCTCTCTCCGTGGGGCCCTCG
 GGGAGGGAGCCCAATGGGGTGTGCGATCCGGCTAAGCTGGAGAGAGGCGACCCGGGAGC

Sp1

 2101 CTGGCGTCCCTGCACCCTGGGAGCGCGAGCGGCGCGGGCGGGGAAGCGCGGCCAGAC
 GACCGCAGGGACGTGGGACCCTCGCGCTCGCCGCGCGCCCCGCCCTTCGCGCCGGGTCTG
 2161 CCCCCGGTCCGCCCGGAGCAGCTGCGCTGTGCGGGCCAGGCCGGGCTCCAGTGGATTGCG
 GGGGCCAGGCGGGCTCGTCGACGCGACAGCCCCGGTCCGGCCCCGAGGGTCACCTAAGC
 2221 CGGGCAACAGACGCCCAGGACCGCGCTTCCACGTGGCGGAGGGACTGGGGACCCGGGCA
 GCCCGTTGTCTGCGGGTCTTGGCGCGAAGGGTGCACCGCCTCCCTGACCCCTGGGCCCCGT

Sp1
 =====
 E2F

 2281 CCGGTCTTCCCCCTTACCTTCCAGCTCCGCTCGTCCGCGCGGAACCCCGCCCCGTCCC
 GGCCAGGACGGGGAAGTGAAGGTGAGGCGGAGCAGGCGCGCTTGGGGCGGGGCAGGG
 2341 GAACCTTCCCGGTCCCCGGCCCCAGCCCCCTTCCGGGCCATCCCAGCCCGTCCCGTTCTT
 CTTGGGAAGGGCCCAGGGGCCGGGTGCGGGAAGGCCCGGTAGGGTCGGGCAGGGCAAGGA

Sp1
 =====

E2F

NFkB

 2401 TTTCCGCGGCCCCGCCCTCTCTCGCGCGCGAGTTTCAGGCAGCGCTGCGTCTGCTGC
 AAAGGCGCCGGGCGGGAGAGGAGCGCCGCGCTCAAAGTCCGTGCGGACGCAGGACGACG

hTRT5'
 *****>

 2461 GCACGTGGGAAGCCCTGGCCCCGGCCACCCCGCGATGCCGCGCGCTCCCCGCTGCCGAG
 CGTGACCCCTTCGGGACCGGGGCCGGTGGGGGCGCTACGGCGCGGAGGGGCGACGGCTC
 2521 CCGTGCGCTCCCTGCTGCGCAGCCACTACCGCGAGGTGCTGCCGCTGGCCACGTTCTGTG
 GGCACGCGAGGGACGACGCGTGGTGATGGCGCTCCACGACGGCGACCGGTGCAAGCAGC

FIG. 21
 (CONTINUED)

2581 GGC GCC TGGGG CCCCAGGGCTGGCGGCTGGTGCAGCGCGGGGACCCGGCGGCTTTCCGCG
CCGCGGACCCCGGGGTCCCCGACCGCCGACCACGTCGCGCCCCCTGGGCCCGCGAAAGGCGC

*

2641 CGCTGGTGGCCAGTGCCTGGTGTGCGTGCCCTGGGACGCACGGCCGCCCCCGCCGCC
GCGACCACCGGGTCACGGACCACACGCACGGGACCCTGCGTGCCGGCGGGGGGCGGCGGG

NFkB

=====

2701 CCTCCTTCCGCCAGGTGGGCCTCCCGGGGTCCGCGTCCGCTGGGGTTGAGGGCGGCCG
GGAGGAAGGCGGTCCACCCGGAGGGGCCCCAGCCGCAGGCCGACCCCAACTCCCCCGCCGC

Topo_II_cleavag

::::::::::::::

NFkB

+++++

NFkB

=====

Intron1

*****>

2761 GGGGGAACCGACGACATGCGGAGAGCAGCGCAGGCGACTCAGGGCGCTTCCCCCGCAGGT
CCCCCTTGGTGCCTGTACGCCTCTCGTCGCGTCCGCTGAGTCCCGGAAGGGGGCGTCCA

e_site

::::

2821 GTCCTGCCTGAAGGAGCTGGTGGCCCGAGTGCCTGCAGAGGCTGTGCGAGCGCGGCGCGAA
CAGGACGGACTTCCTCGACCACCGGGCTCACGACGCTCTCCGACACGCTCGCGCCCGCGCTT

2881 GAACGTGCTGGCCTTCGGCTTCGCGCTGCTGGACGGGGCCCGCGGGGGCCCCCGAGGC
CTTGACGACCGGAAGCCGAAGCGCGACGACCTGCCCCGGGCGCCCCCGGGGGGCTCCG

2941 CTTACACCACGCGTGCGCAGCTACCTGCCCCAACCGGTGACCGACGCACTGCGGGGGAG
GAAGTGGTGGTGCACGCGTCGATGGACGGGTTGTGCCACTGGCTGCGTGACCCCCCTC

3001 CGGGGCGTGGGGGCTGCTGCTGCGCCGCGTGGGCGACGACGTGCTGGTTACCTGCTGGC
GCCCCGCACCCCGACGACGACGCGCGCACCCGCTGCTGCACGACCAAGTGGACGACCG

3061 ACGCTGCGCGCTCTTTGTGCTGGTGGCTCCAGCTGCGCCTACCAGGTGTGCGGGCCGCC
TGCGACGCGCGAGAAACACGACCACCGAGGGTCGACGCGGATGGTCCACACGCCCGGCGG

3121 GCTGTACCAGCTCGGCGCTGCCACTCAGGCCCGGGCCCCCGCCACACGCTAGTGGACCCCG
CGACATGGTTCGAGCCGCGACGGTGAGTCCGGGCGGGGGCGGTGTGCGATCACCTGGGGC

3181 AAGGCGTCTGGGATGCGAACGGGCTGGAACCATAGCGTCAGGGAGGCCGGGGTCCCCCT
TTCCGACAGACCTACGCTTGCCCGGACCTTGGTATCGCAGTCCCTCCGGCCCCAGGGGA

3241 GGGCCTGCCAGCCCCGGGTGCGAGGAGGCGCGGGGGCAGTGCCAGCCGAAGTCTGCCGTT
CCCGGACGGTGGGGGCCACGCTCCTCCGCGCCCCCGTACGGTTCGGCTTCAGACGGCAA

3301 GCCCAAGAGGCCCAGGCGTGGCGCTGCCCTGAGCCGGAGCGGACGCCCCGTGGGCAGGG
CGGGTTCTCCGGGTCCGCACCGCGACGGGACTCGGCCTCGCCTGCGGGCAACCCGTCCC

FIG. 21

(CONTINUED)

3361 GTCCTGGGCCCCACCCGGGCAGGACGCGTGGACCGAGTGACCGTGGTTTCTGTGTGGTGTG
 CAGGACCCGGGTGGGCCCCGTCTGCGCACCTGGCTCACTGGCACCAAAGACACACCACAG
 3421 ACCTGCCAGACCCGCCGAAGAAGCCACCTCTTTGGAGGGTGCCTCTCTGGCACGCGCCA
 TGGACGGTCTGGGCGGCTTCTTCGGTGGAGAAACCTCCACGCGAGAGACCGTGCCTGGT
 3481 CTCCACCCATCCGTGGGCGGCCAGCACACGCGGGCCCCCATCCACATCGCGGCCACC
 GAGGGTGGGTAGGCACCCGGCGGTCTGTGGTGCCTCCGGGGGTAGGTGTAGCGCCGGTGG
 3541 ACGTCCCTGGGACACGCCTTGTCCCCCGGTGTACGCCGAGACCAAGCACTTCCTCTACTC
 TGCAGGGACCCCTGTGCGGAACAGGGGGCCACATGCGGCTCTGGTTCGTGAAGGAGATGAG
 3601 CTCAGGCGACAAGGAGCAGCTGCGGCCCTCCTTCCTACTCAGCTCTCTGAGGCCCAGCCT
 GAGTCCGCTGTTCCTCGTCGACGCGGGAGGAAGGATGAG'CGAGAGACTCCGGGTCCGA
 3661 GACTGGCGCTCGGAGGCTCGTGGAGACCATCTTTCTGGGTTCAGGCCCTGGATGCCAGG
 CTGACCGCGAGCCTCCGAGCACCTCTGGTAGAAAGACCCAAGGTCCGGGACCTACGGTCC
 3721 GACTCCCCGCGAGGTTGCCCCGCTGCCCGAGCGCTACTGGCAAATGCGGCCCTGTTTCT
 CTGAGGGGCGTCCAACGGGGCGGACGGGTGCGGATGACCGTTTACGCCGGGGACAAAGA
 3781 GGAGCTGCTTGGGAACCACGCGCAGTGCCCTACGGGGTGCTCCTCAAGACGCACTGCCC
 CCTCGACGAACCTTGGTGCCTGTCACGGGGATGCCCCACGAGGAGTTCGTGCTGACGGG
 3841 GCTGCGAGCTGCGGTACCCCGAGCAGCCGGTGTCTGTGCCCCGGGAGAAGCCCCAGGGCTC
 CGACGCTCGACGCCAGTGGGGTCTGCGCCACAGACACGGGCCCTCTTCGGGGTCCCGAG
 3901 TGTGGCGGCCCCCGAGGAGGAGGACACAGACCCCCGTGCTGCTGGTGCAGCTGCTCCGCCA
 ACACCGCCGGGGGCTCCTCCTCCTGTGTCTGGGGGCGAGCGGACCACGTCGACGAGGCGGT
 3961 GCACAGCAGCCCCTGGCAGGTGTACGGCTTCGTGCGGGCCTGCCTGCGCCGGCTGGTGCC
 CGTGTCTGCGGGGACCGTCCACATGCCGAAGCACGCCCCGACGGACGCGGCCGACACGG
 4021 CCCAGGCCTCTGGGGCTCCAGGCACAACGAACGCCGCTTCCTCAGGAACACCAAGAAGTT
 GGGTCCGGAGACCCCGAGGTCCGTGTTGCTTGCCTGCGGCGAAGGAGTCCTTGTGGTTCTTCAA
 4081 CATCTCCCTGGGGAAGCATGCCAAGCTCTCGCTGCAGGAGCTGACGTGGAAGATGAGCGT
 GTAGAGGGACCCCTTCGTACGGTTCGAGAGCGACGTCCTCGACTGCACCTTCTACTCGCA

 4141 GCGGGACTGCGCTTGGCTGCGCAGGAGCCAGGTGAGGAGGTGGTGGCCGTCGAGGGCCC
 CGCCCTGACGCGAACCGACGCGTCTCGGGTCCACTCCTCCACCACCGGCAGTCCCGGG
 Intron2

 4201 AGGCCCCAGAGCTGAATGCAGTAGGGGCTCAGAAAAGGGGGCAGGCAGAGCCCTGGTCTT
 TCCGGGGTCTCGACTTACGTATCCCCGAGTCTTTTCCCCCGTCCGTCTCGGGACCAGGA

 4261 CCTGTCTCCATCGTCACGTGGGCACACGTGGCTTTTCGCTCAGGACGTGAGTGACACG
 GGACAGAGGTAGCAGTGCACCCGTGTGCACCGAAAAGCGAGTCCTGCAGCTCACCTGTGC
 *****>
 4321 GTGATCGAGGTGAC
 CACTAGCTCCAGCTG

FIG. 21
(CONTINUED)

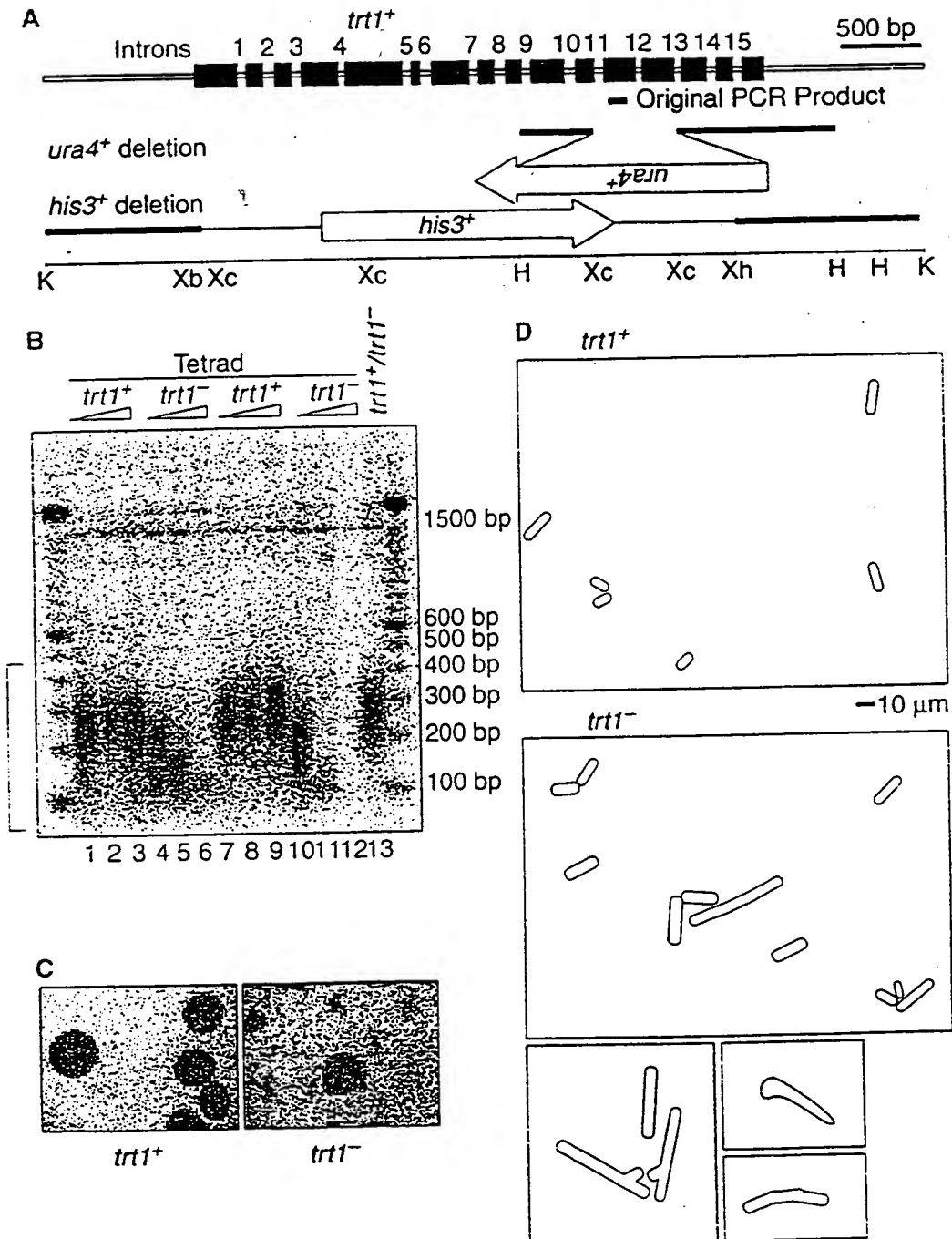


FIG. 22

gccaagttcctgcactggctgatgagtgtgtacgtcgctcgagctgctcaggtctttcttt
 tatgtcacggagaccacgtttcaaaagaacaggctctttttctaccggaagagtgtctgg
 agcaagttgcaaagcattggaatcagacagcacttgaagaggggtgcagctgcgggacgtg
 tcggaagcagaggtcaggcagcatcggaagccaggcccgccctgctgacgtccagactc
 cgcttcaccccaagcctgacgggctgcggccgattgtgaacatggactacgtcgtggga
 gccagaacgttccgcagagaaaagagggccgagcgtctcacctcgaggggtgaaggcactg
 ttcagcgtgctcaactacgagcgggcgcg

FIG. 23

TCTACCTTGACAGACCTCCAGCCGTACATGCGACAGTTCGTGGCTCACCTGCAGGAG
 ACCAGCCCGCTGAGGGATGCCGTCGTCATCGAGCAGAGCTCCTCCCTGAATGAGGCC
 AGCAGTGGCCTCTTCGACGTCTTCCTACGCTTCATGTGCCACCAACGCCGTGCGCATC
 AGGGGCAAGTC

FIG. 24

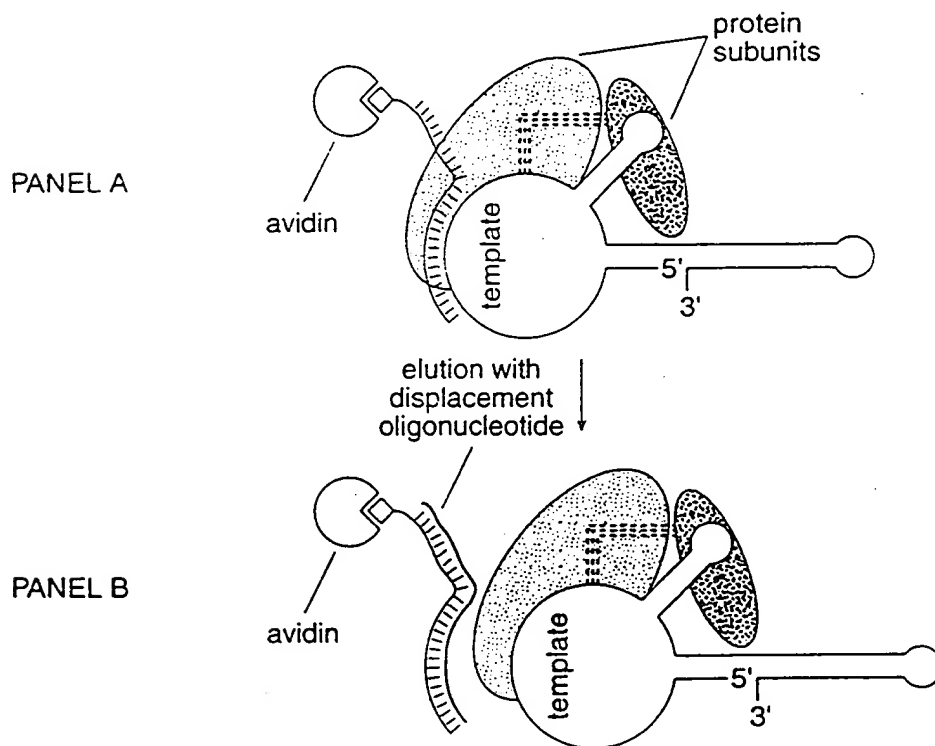


FIG. 26

35/103



FIG. 25

36/103

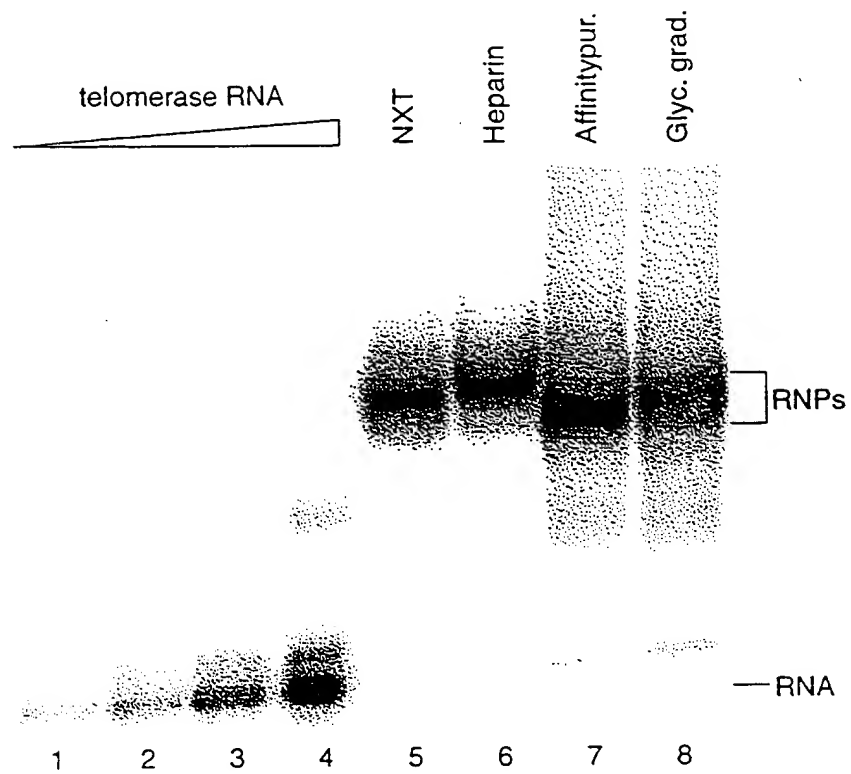


FIG. 27

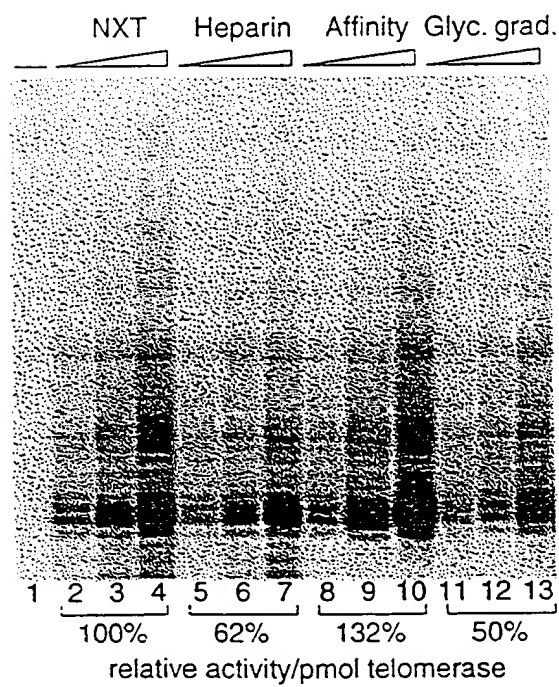


FIG. 28

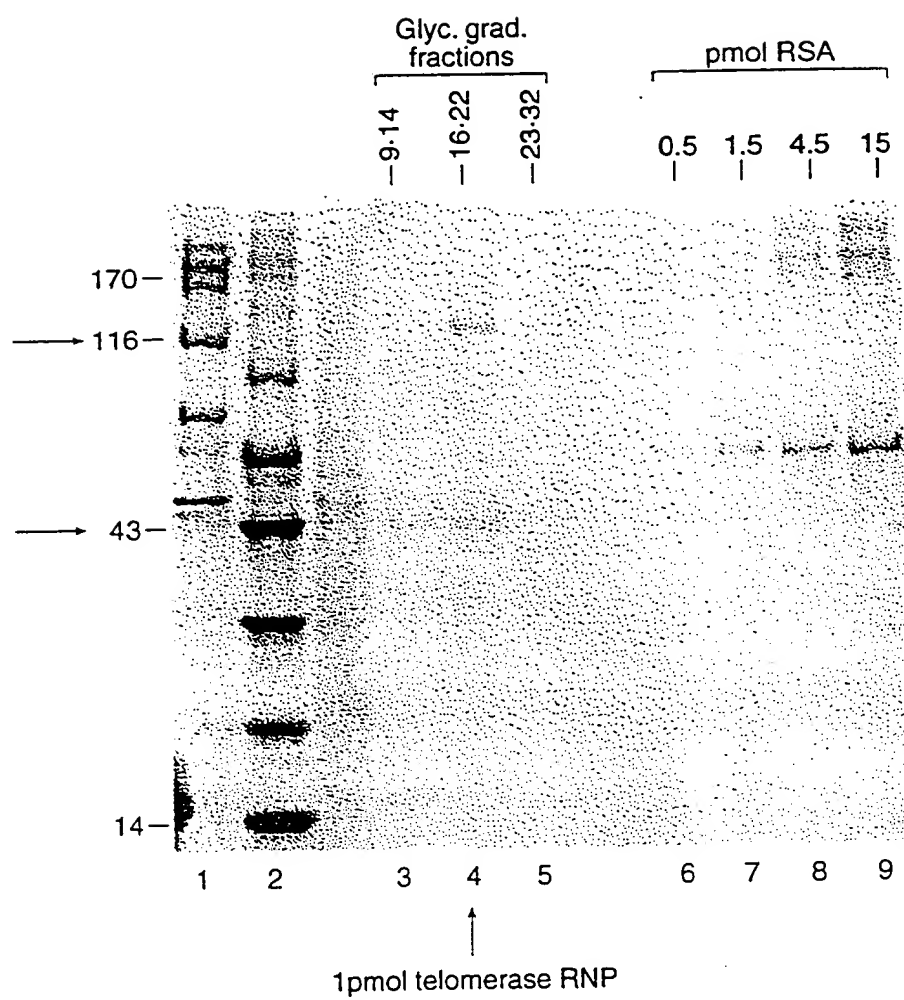


FIG. 29

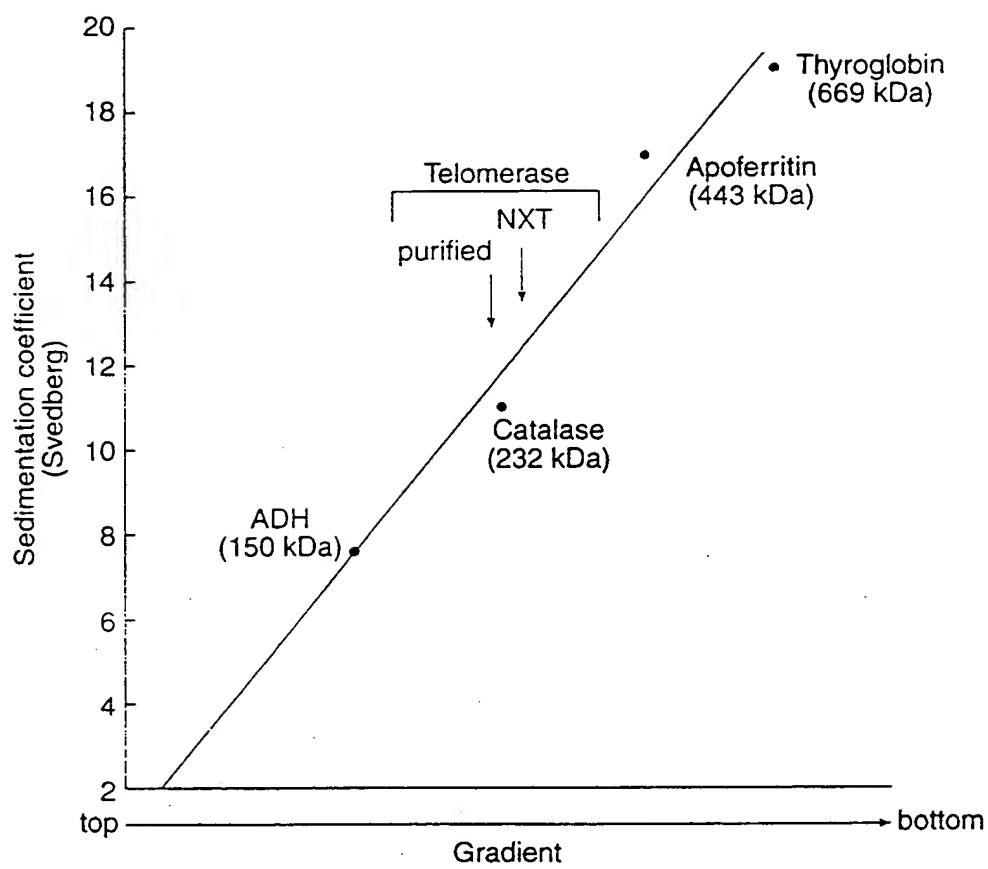


FIG. 30

Telomerase:

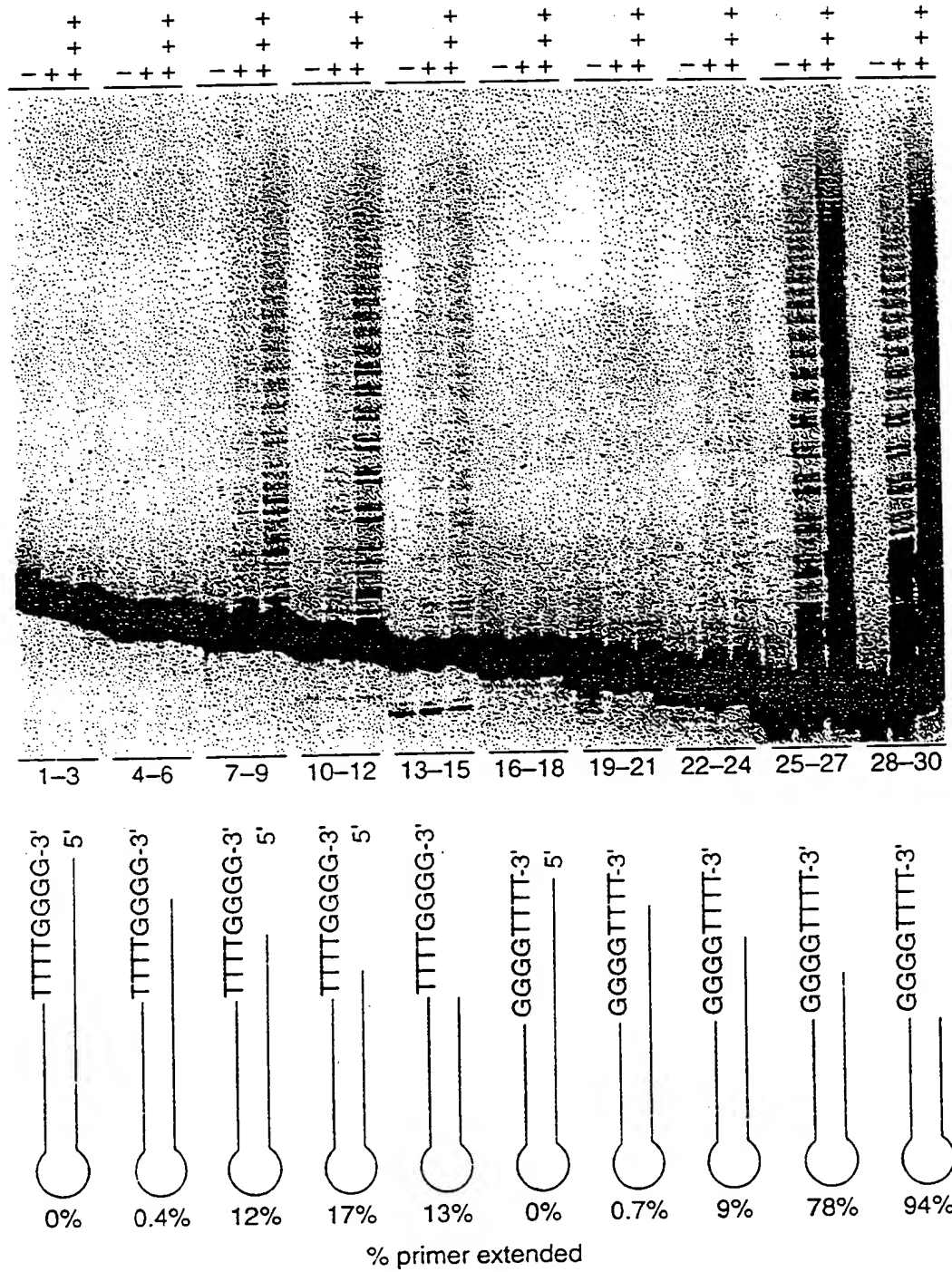


FIG. 31

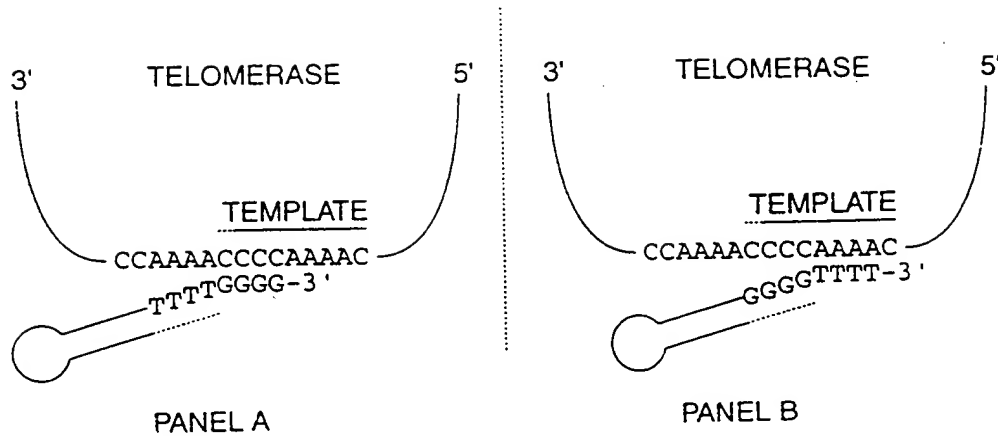


FIG. 32

1	CCCCAAAACC	CCAAAACCCC	AAAACCCCTA	TAAAAAAGA	AAAAATTGAG
51	GTAGTTTAGA	AATAAAATAT	TATTCCTGCA	CAAATGGAGA	TGGATATTGA
101	TTTGGATGAT	ATAGAAAATT	TACTTCCTAA	TACATTCAAC	AAGTATAGCA
151	GCTCTTGTAG	TGACAAGAAA	GGATGCAAAA	CATTGAAATC	TGGCTCGAAA
201	TCGCCTTCAT	TGACTATTCC	AAAGTTGCAA	AAACAATTAG	AGTTCTACTT
251	CTCGGATGCA	AATCTTTATA	ACGATTCTTT	CTTGAGAAAA	TTAGTTTTAA
301	AAAGCGGAGA	GCAAAGAGTA	GAAATTGAAA	CATTACTAAT	GTTTAAATAA
351	AATCAGGTAA	TGAGGATTAT	TCTATTTTTT	AGATCACTTC	TTAAGGAGCA
401	TTATGGAGAA	AATTACTTAA	TACTAAAAGG	TAAACAGTTT	GGATTATTTT
451	CCTAGCCAAC	AATGATGAGT	ATATTAAATT	CATATGAGAA	TGAGTCAAAG
501	GATCTCGATA	CATCAGACTT	ACCAAAGACA	AACTCGCTAT	AAAACGCAAG
551	AAAAAGTTTG	ATAATCGAAC	AGCAGAAGAA	CTTATTGCAT	TTACTATTTC
601	TATGGGTTTT	ATTACAATTG	TTTTAGGTAT	CGACGGTGAA	CTCCCGAGTC
651	TTGAGACAAT	TGAAAAAGCT	GTTTACAAC	GAAGGAATCG	CAGTTCTGAA
701	AGTTCTGATG	TGTATGCCAT	TATTTTGTGA	ATTAATCTCA	AATATCTTAT
751	CTCAATTTAA	TGGATAGCTA	TAGAAACAAA	CCAAATAAAC	CATGCAAGTT
801	TAATGGAATA	TACGTTAAAT	CCTTTGGGAC	AAATGCACAC	TGAATTTATA
851	TTGGATTCTT	AAAGCATAGA	TACACAGAAT	GCTTTAGAGA	CTGATTTAGC
901	TTACAACAGA	TTACCTGTTT	TGATTACTCT	TGCTCATCTC	TTATATCTTT
951	AAAAGAAGCA	GGCGAAATGA	AAAGAAGACT	AAAGAAAGAG	ATTTCAAAAT
1001	TTGTTGATT	TTCTGTAACC	GGAATTAACA	ACAAGAATAT	TAGCAACGAA
1051	AAAGAAGAAG	AGCTATCACA	ATCCTGATT	TTAAAGATTT	CAAAAATTCC
1101	AGGTAAGAGA	GATACATTCA	TTAAAATTCA	TATATTATAG	TTTTTCATTT
1151	CACAGCTGTT	ATTTTCTTTT	ATCTTAACAA	TATTTTTTGA	TTAGCTGGAA
1201	GTA AAAAGTA	TCAAATAAGA	GAAGCGCTAG	ACTGAGGTAA	CTTAGCTTAT
1251	TCACATTCAT	AGATCGACCT	TCATATATCC	AATACGATGA	TAAGGAAACA
1301	GCAGTCATCC	GTTTTTAAAA	TAGTGCTATG	AGGACTAAAT	TTT TAGAGTC
1351	AAGAAATGGA	GCCGAAATCT	TAATCAAAAA	GAATTGCGTC	GATATTGCAA
1401	AAGAATCGAA	CTCTAAATCT	TTCGTTAATA	AGTATTACCA	ATCTTGATTG
1451	ATTGAAGAGA	TTGACGAGGC	AACTGCACAG	AAGATCATTA	AAGAAATAAA
1501	GTAACTTTTA	TTAATTAGAG	AATAAACTAA	ATTACTAATA	TAGAGATCAG
1551	CGATCTTCAA	TTGACGAAAT	AAAAGCTGAA	CTAAAGTTAG	ACAATAAAAA
1601	ATACAAACCT	TGGTCAAAAT	ATTGAGGAAG	GAAAAGAAGA	CCAGTTAGCA
1651	AAAGAAAAAA	TAAGGCAATA	AATAAAATGA	GTACAGAAGT	GAAGAAATAA
1701	AAGATTTATT	TTTTTCAATA	ATTTATTGAA	AAGAGGGGTT	TTGGGGTTTT
1751	GGGGTTTTGG	GG			

FIG. 34

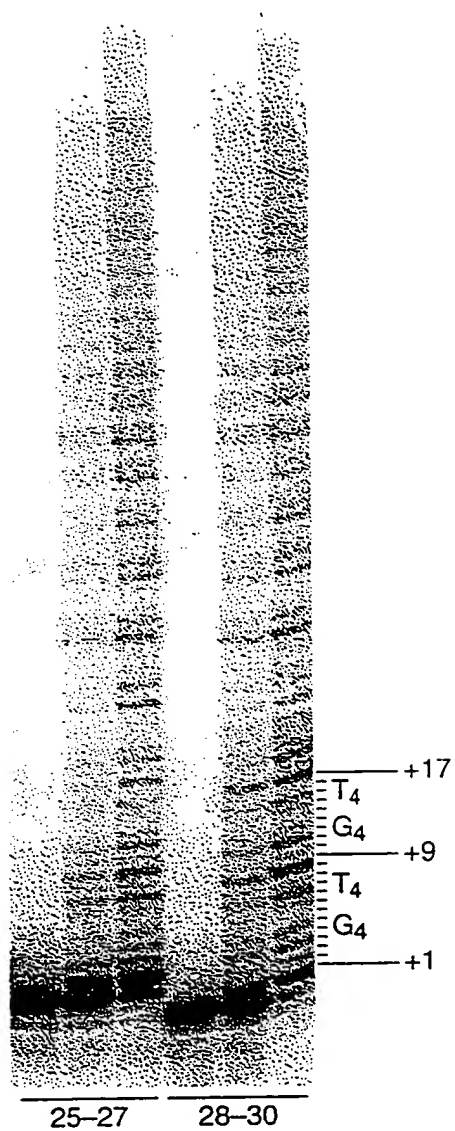


FIG. 33

1 CCCCCAAAACCCAAAACCCCAAAACCCCTATAAAAAAAGAAAAAATTGAGGTAGTTTAGA 60
GGGGTTTTGGGGTTTTGGGGTTTTGGGGATATTTTTTTCTTTTTTAACCTCCATCAAATCT

a P Q N P K T P K P L * K K K K L R * F R -
b P K T P K P Q N P Y K K R K N * G S L E -
c P K P Q N P K T P I K K E K I E V V * K -

61 AATAAAATATTATTCCCGCACAAATGGAGATGGATATTGATTGGATGATATAGAAAATT 120
TTATTTTATAATAAGGGCGTGTTCCTCTACCTATAACTAAACCTACTATATCTTTTAA

a N K I L F P H K W R W I L I W M I * K I -
b I K Y Y S R T N G D G Y * F G * Y R K F -
c * N I I P A Q M E M D I D L D D I E N L -

121 TACTTCCTAATACATTCAACAAGTATAGCAGCTCTTGTAAGTGACAAGAAAGGATGCAAAA 180
ATGAAGGATTATGTAAGTTGTTTCATATCGTCGAGAACATCACTGTTCTTTCTACGTTTT

a Y F L I H S T S I A A L V V T R K D A K -
b T S * Y I Q Q V * Q L L * * Q E R M Q N -
c L P N T F N K Y S S S C S D K K G C K T -

181 CATTGAAATCTGGCTCGAAATCGCCTTCATTGACTATTCCAAAGTTGCAAAAACAATTAG 240
GTAACTTTAGACCGAGCTTTAGCGGAAGTAACTGATAAGGTTTCAACGTTTTTGTTAATC

a H * N L A R N R L H * L F Q S C K N N * -
b I E I W L E I A F I D Y S K V A K T I R -
c L K S G S K S P S L T I P K L Q K Q L E -

241 AGTTCTACTTCTCGGATGCAAATCTTTATAACGATTCTTTCTTGAGAAAATTAGTTTTAA 300
TCAAGATGAAGAGCCTACGTTTAGAAATATTGCTAAGAAAGAACTCTTTTAATCAAATT

a S S T S R M Q I F I T I L S * E N * F * -
b V L L L G C K S L * R F F L E K I S F K -
c F Y F S D A N L Y N D S F L R K L V L K -

301 AAAGCGGAGAGCAAAGAGTAGAAATTGAAACATTACTAATGTTTAAATAAAATCAGGTAA 360
TTTCGCCTCTCGTTTCTCATCTTTAACTTTGTAATGATTACAAATTTATTTTAGTCCATT

a K A E S K E * K L K H Y * C L N K I R * -
b K R R A K S R N * N I T N V * I K S G N -
c S G E Q R V E I E T L L M F K * N Q V M -

361 TGAGGATTATTCTATTTTTTAGATCACTTCTTAAGGAGCATTATGGAGAAAATTACTTAA 420
ACTCCTAATAAGATAAAAAATCTAGTGAAGAATTCCTCGTAATACCTCTTTTAATGAATT

a * G L F Y F L D H F L R S I M E K I T * -
b E D Y S I F * I T S * G A L W R K L L N -
c R I I L F F R S L L K E H Y G E N Y L I -

TACTAAAAGGTAAACAGTTTGGATTATTTCCCTAGCCAACAATGATGAGTATATTAATT
-----+-----+-----+-----+-----+ 480

a Y * K V N S L D Y F P S Q Q * * V Y * I -
b T K R * T V W I I S L A N N D E Y I K F -
c L K G K Q F G L F P * P T M M S I L N S -

CATATGAGAATGAGTCAAAGGATCTCGATACATCAGACTTACC AAAGACAAACTCGCTAT
-----+-----+-----+-----+-----+ 540

a H M R M S Q R I S I H Q T Y Q R Q T R Y -
b I * E * V K G S R Y I R L T K D K L A I -
c Y E N E S K D L D T S D L P K T N S L * -

AAAA CGCAAGAAAAAGTTTGATAATCGAACAGCAGAAGAACTTATTGCATT TACTATTTCG
-----+-----+-----+-----+-----+ 600

a K T Q E K V * * S N S R R T Y C I Y Y S -
b K R K K K F D N R T A E E L I A F T I R -
c N A R K S L I I E Q Q K N L L H L L F V -

TATGGGTTTTATTACAATTGTTTTAGGTATCGACGGTGA ACTCCCAGTCTTGAGACAAT
-----+-----+-----+-----+-----+ 660

a Y G F Y Y N C F R Y R R * T P E S * D N -
b M G F I T I V L G I D G E L P S L E T I -
c W V L L Q L F * V S T V N S R V L R Q L -

TGAAAAAGCTGTTTACA ACTGAAGGAATCGCAGTTCTGAAAGTTCTGATGTGTATGCCAT
-----+-----+-----+-----+-----+ 720

a * K S C L Q L K E S Q F * K F * C V C H -
b E K A V Y N * R N R S S E S S D V Y A I -
c K K L F T T E G I A V L K V L M C M P L -

TATTTTGTGAATTAATCTCAA ATATCTTATCTCAATTTAATGGATAGCTATAGAAACAA
-----+-----+-----+-----+-----+ 780

a Y F V N * S Q I S Y L N L M D S Y R N K -
b I L * I N L K Y L I S I * W I A I E T N -
c F C E L I S N I L S Q F N G * L * K Q T -

CCAAATAAACCATGCAAGTTTAATGGAATATAC GTTAAATCCTTTGGGACAAATGCACAC
-----+-----+-----+-----+-----+ 840

a P N K P C K F N G I Y V K S F G T N A H -
b Q I N H A S L M E Y T L N P L G Q M H T -
c K * T M Q V * W N I R * I L W D K C T L -

TGAATTTATATTGGATTCTTAAAGCATAGATACACAGAAT GCCTTTAGAGACTGATTTAGC
-----+-----+-----+-----+-----+ 900

a * I Y I G F L K H R Y T E C F R D * F S -
b E F I L D S * S I D T Q N A L E T D L A -
c N L Y W I L K A * I H R M L * R L I * L -

```

TTACAACAGATTACCTGTTTTGATTACTCTTGCTCATCTCTTATATCTTTAAAAGAAGCA
901 -----+-----+-----+-----+-----+-----+-----+-----+-----+ 960
AATGTTGTCTAATGGACAAAATAATGAGAACGAGTAGAGAATATAGAAATTTTCTTCGT

a L Q Q I T C F D Y S C S S L I S L K E A -
b Y N R L P V L I T L A H L L Y L * K K Q -
c T T D Y L F * L L L L I S Y I F K R S R -

GGCGAAATGAAAAGAAGACTAAAGAAAGAGATTTCAAATTTGTTGATTCTTCTGTAACC
961 -----+-----+-----+-----+-----+-----+-----+-----+-----+ 1020
CCGCTTTACTTTTCTTCTGATTTCTTTCTCTAAAGTTTTAAACAATAAGAAGACATTGG

a G E M K R R L K K E I S K F V D S S V T -
b A K * K E D * R K R F Q N L L I L L * P -
c R N E K K T K E R D F K I C * F F C N R -

GGAATTAACAACAAGAATATTAGCAACGAAAAAGAAGAAGAGCTATCACAATCCTGATTC
1021 -----+-----+-----+-----+-----+-----+-----+-----+-----+ 1080
CCTTAATTGTTGTTCTTATAATCGTTGCTTTTTCTTCTCTCGATAGTGTAGGACTAAG

a G I N N K N I S N E K E E E L S Q S * F -
b E L T T R I L A T K K K K S Y H N P D S -
c N * Q Q E Y * Q R K R R R A I T I L I L -

TTAAAGATTTCAAATTCAGGTAAGAGAGATACATTCAATTAATTCATATATTATAG
1081 -----+-----+-----+-----+-----+-----+-----+-----+-----+ 1140
AATTTCTAAAGTTTTTAAGGTCCATTCTCTCTATGTAAGTAATTTTAAGTATATAATATC

a L K I S K I P G K R D T F I K I H I L * -
b * R F Q K F Q V R E I H S L K F I Y Y S -
c K D F K N S R * E R Y I H * N S Y I I V -

TTTTTCATTTACAGCTGTTATTTCTTTTATCTTAACAATATTTTTTGATTAGCTGGAA
1141 -----+-----+-----+-----+-----+-----+-----+-----+-----+ 1200
AAAAAGTAAAGTGTGACAATAAAAGAAAATAGAATTGTTATAAAAAACTAATCGACCTT

a F F I S Q L L F S F I L T I F F D * L E -
b F S F H S C Y F L L S * Q Y F L I S W K -
c F H F T A V I F F Y L N N I F * L A G S -

GTAAAAAGTATCAAATAAGAGAAGCGCTAGACTGAGGTAAGTATTGCTTATTCACATTCAT
1201 -----+-----+-----+-----+-----+-----+-----+-----+-----+ 1260
CATTTTTTCATAGTTTATTCTCTTCGCGATCTGACTCCATTGAATCGAATAAGTGTAGTA

a V K S I K * E K R * T E V T * L I H I H -
b * K V S N K R S A R L R * L S L F T F I -
c K K Y Q I R E A L D * G N L A Y S H S * -

AGATCGACCTTCATATATCCAATACGATGATAAGGAAACAGCAGTCATCCGTTTTAAAAA
1261 -----+-----+-----+-----+-----+-----+-----+-----+-----+ 1320
TCTAGCTGGAAGTATATAGGTTATGCTACTATTCTTTGTCGTCAGTAGGCAAAATTTT

a R S T F I Y P I R * * G N S S H P F * K -
b D R P S Y I Q Y D D K E T A V I R F K N -
c I D L H I S N T M I R K Q Q S S V L K I -

TAGTGCTATGAGGACTAAATTTTTAGAGTCAAGAAATGGAGCCGAAATCTTAATCAAAAA
1321 -----+-----+-----+-----+-----+-----+-----+-----+-----+ 1380
ATCAGGATACTCCTGATTTAAAAATCTCAGTTCTTTACCTCGGCTTTAGAATTAGTTTTT

a * C Y E D * I F R V K K W S R N L N Q K -
b S A M R T K F L E S R N G A E I L I K K -
c V L * G L N F * S Q E M E P K S * S K R -

```

FIG. 35
(CONTINUED)

1381 GAATTGCGTCGATATTGCAAAAGAATCGAACTCTAAATCTTTTCGTTAATAAGTATTACCA
 -----+-----+-----+-----+-----+-----+-----+ 1440
 CTTAACGCAGCTATAACGTTTTCTTAGCTTGAGATTAGAAAAGCAATTATTCATAATGGT

 a E L R R Y C K R I E L * I F R * * V L P -
 b N C V D I A K E S N S K S F V N K Y Y Q -
 c I A S I L Q K N R T L N L S L I S I T N -

 ATCTTGATTGATTGAAGAGATTGACGAGGCAACTGCACAGAAGATCATTAAAGAAATAAA
 1441 -----+-----+-----+-----+-----+ 1500
 TAGAACTAACTAATTCTCTAACTGCTCCGTTGACGTGTCTTCTAGTAATTTCTTTATTT

 a I L I D C R D * R G N C T E D H * R N K -
 b S * L I E E I D E A T A Q K I I K E I K -
 c L D * L K R L T R Q L H R R S L K K * S -

 GTAACCTTTTATTAATTAGAGAATAAACTAAATTACTAATATAGAGATCAGCGATCTTCAA
 1501 -----+-----+-----+-----+-----+ 1560
 CATTGAAAATAATTAATCTCTTATTTGATTTAATGATTATATCTCTAGTCGCTAGAAGTT

 a V T F I N * R I N * I T N I E I S D L Q -
 b * L L L I R E * T K L L I * R S A I F N -
 c N F Y * L E N K L N Y * Y R D Q R S S I -

 TTGACGAAATAAAAGCTGAACTAAAGTTAGACAATAAAAAATACAAACCTTGGTCAAAAT
 1561 -----+-----+-----+-----+-----+ 1620
 AACTGCTTTATTTTCGACTTGATTTCAATCTGTTATTTTTTATGTTTGAACCAGTTTAA

 a L T K * K L N * S * T I K N T N L G Q N -
 b * R N K S * T K V R Q * K I Q T L V K I -
 c D E I K A E L K L D N K K Y K P W S K Y -

 ATTGAGGAAGGAAAAGAAGACCAGTTAGCAAAAGAAAAATAAGGCAATAAATAAAATGA
 1621 -----+-----+-----+-----+-----+ 1680
 TAACTCCTTCCTTTTCTTCTGGTCAATCGTTTTCTTTTTTATTCGGTTATTTATTTTACT

 a I E E G K E D Q L A K E K I R Q * I K * -
 b L R K E K K T S * Q K K K * G N K * N E -
 c * G R K R R P V S K R K N K A I N K M S -

 GTACAGAAGTGAAGAAATAAAAGATTTATTTTTTTCAATAATTATTGAAAAGAGGGGTT
 1681 -----+-----+-----+-----+-----+ 1740
 CATGTCTTCACTTCTTTATTTTCTAAATAAAAAAAGTTATTAAATAACTTTTCTCCCCAA

 a V Q K * R N K R F I F F N N L L K R G V -
 b Y R S E E I K D L F F S I I Y * K E G F -
 c T E V K K * K I Y F F Q * F I E K R G F -

 TTGGGGTTTTTGGGGTTTTTGGGG
 1741 -----+-----+-----+-----+ 1762
 AACCCCAAAACCCCAAAACCCC

 a L G F W G F G -
 b W G F G V L G -
 c G V L G F W -

FIG. 35
(CONTINUED)

[illegible]

```

798 EKLINVSRENGSFKFNMMK.LQTSFPLSPSKFAKYGMSVEEQNIVQDYCD 846
      .....|.|||..|:|:|.....|...|
577 .....PGDELRPSMQLLQEKGKLGGLTDFPYECIDEWTKNKTHVD 617
847 WIGISIDMKTLALMPNINLRIEGILCTNLNMOTKKASMWLKKKKLSFLM 896
      |.| |||.. . :|:|:| ..|:|. :
618 NIVILSDMMIAEGYS DINVRGSSIVNSI.....KKYKDEVN 653
897 NNITHYFRKTITTEDFANKTLNKLFISSGGYKYMCAKEYKD.HFKKNLAM 945
      ||. : ..|:|:| |:|:|:|:|:|:|
654 PNIKIF...AVDLEGYG.....KCLNLGDEFNNNYIKIFGM 687
946 SSMIDLEVSKIISVTRAFFKYLVCNIKDTIFGEEHYDPDFLSTLKHFIE 995
      |.| .....|:|:| .. : :|
688 SDSI.....LKFISAKQGGA.....NMVE 706
996 IFSTKKYIFNRVC 1008
      :: |.: :.:
707 VI..KNFALQKIG 717

```

FIG. 36
(CONTINUED)

132	LSLTKQKYFFQDEWNQVRAMIGNEL.FRHLYTKYLIFQRTSE..GTLVQFC	178
1	MSRRNQ.....KKPQAPIGNETNLDFVLQNLLEVYKSQIEHYKTQQQQI	43
179	GNNVFDHLKVNDKFDKKQKGGAAADMNEPRCCSTCKYNVKNKEDHFLNNIN	228
44	KEEDLKLLKFKNQDQDGNSSGNDDEE.....NNSNKQQELLRRVN	84
229	VPNWNMKSRTTRIFYCTHFNRRNQFFKKHEFVSNKNNISAMDRAQTIFTN	278
85QIKQVQLIKK...VGSKVEKDLNLNEDENKKN	114
279	IFRFNRIRKKLKDVKIEKIAYMLEKVKDFNFNYLTKSCPLPENWRERKQ	328
115	GLSEQQVKEEQLRTITEEQVKYQNLVFNMDYQLDLNESGGHRRHRRET DY	164
329	KIENLINKTREESKYEEELFSYTTDNKCVTQFINE.FFYNILPKDFLTG	377
165	DTEKWFEISHDQK.....NYVSIYANQKTSYCWWLKDYFNK	200
378	RNRKNFQKKVKKYVELNKHელიHKNLLEKINTREISWMQVETSAKHFFY	427
201	NNYDHLNVSINRLE..TEAEFYAFDDFSQTIKLTNNSYQTVNID.....	242
428	FDHENIYVLWKLLRWI..FEDLVVSLIRCFYVTEQQKSYSKTYYYRKNI	475
243	VNFDNNLCILALLRFLSLERFNILNIRSSY..TRNQYNFEKIGELLETI	290
476	WDVIMKMSIADLKKETLAEVQEKEVEEWKKS LGFAPGKLRLIPKKTFRP	525
291	FAVVF SHR.....HLQGIHLQVPCEAFQYLVNSSSQISVKDSQLQ	330
526	IMTFNKKIVNSDRKTTKLTNTKLLNSHMLKTLKNRMFKDPFGFAVFNY	575
331	VYSFSTDLKLVD..TNKVQDYFKFLQEFPRLTHVSQQAIPVSATNAVENL	378

FIG. 37

576 DDVMKKYEEFVCKWKQVGPQLF.FATMDIEKCYDS..VNREK 615
 : : : | | : : | | : : : | : : : | : : : |
 379 NVLLKKVKH ANLNLVSIPTQFNFDYFVNQLHLKLEFGLEPNILTKQK 426
 516 LSTFL.....KTTKLLSSDFWIMTAQILKRKNNI..VIDSKNFRKKEMK 657
 | : : | | : : | | : : | | : : | | : : | | : : |
 427 LENLLLSIKQSKNLKFLRLNFYTYVAQETSRKQILKQATTIKNLKNNKQ 476
 558 DYFRQKFQKIALEGGQYPTLFSVLEN..EQNDLNAKKT LIVEAKQRNYFK 705
 : : : | : : | : : | : : | : : | : : | : : |
 477 EETPETKDETPSESTSGMKFFDHLSELTELEDFSVN....LQATQEY.. 520
 706 KDNLLQPVINICQYNYINFNGKFYKQTKGIPQGLCVSSILSSFYYATLEE 755
 | : : | | : : | | : : | | : : | | : : | | : : |
 521 .DSLHKLLIRSTNLKKFKLSYKYEMEKSKMDTFIDLKNI.....YETLNN 564
 756 SSLGFLRDESMNPENPNVNLLMRLTDDYLLITTQENNAVL FIEKLINVS 305
 : : : | : : | : : | : : | : : | : : | : : |
 565LKRCVNI SNPHGNISYELTN.....KDSTFYKFKLTLNQE 500
 806 ENGFKFNMKKLQTSFPLSPSKFAKYGMDSVEEQNIVQDYCDWIGISIDMK 855
 | : : | | : : | | : : | | : : | | : : | | : : |
 601 LQHAKYTFK..QNEFQFN NVKSAKIESSLESLEDIDSLCKSIASCKNLQ 648
 856 TLALMPNINLRIEGILCTLNLMQT..KKASMWLKK..KLKSFLMNNITH 901
 : : : | : : | : : | : : | : : | : : | : : |
 649 NVNI.....IASLLYPNNIQKNP FNKPNLLFFKQFEQLKNLE NVSINC 691
 902 YFRKTI...TTEDFANKTLNKLFISSGGYKYMQCAKEYKDHFKKNLAMSSM 948
 : : : | : : | : : | : : | : : | : : | : : |
 692 ILDQHILNSISEFLEKNKKIKAFILKRYLLQYYLDYTKLFKTLQQLPEL 741
 949 IDLEVSKIIYSVT.....RAFFKYLVCNIKDT..IFGEEHY 982
 : : : | : : | : : | : : | : : | : : | : : |
 742 NQVYINQQLEELTVSEVHKQVWENHKQKAFYEPLCEFIKESSQTLQLIDF 791
 983 PDFFLS TLKHFIEIFSTKKY IFNRVCMILKAKEAKLKSDDQCQSLIQ 1028
 : : : | : : | : : | : : | : : | : : | : : |
 792 DQNTVSDDSIKKILESISESKYHHYLRNLNPSQSSSLIKSENEEIQELLK 840

FIG. 37
(CONTINUED)

4 DIDLDDIENLLPNTFNKYSSSCSDKKGCKTLKSGSKSPSLTIPK..... 47
 : : : | | : : : | | : : | | : : | : : |
 617 NVKSAKIESSLESLEDIDSLCKSIASCKNLQNVNIIASLLYPNNIQKNP 666
 48LQKQLEFYFSDANLYNDSFLRKLVLKSGEQRVE....IETLLM 86
 : : : | : : | : : | : : | : : | : : | : : |
 667 FNKPNLLFFKQFEQLKNLE NVSINCILDQHILNSISEFLEKNKKIKAFIL 716

FIG. 38

49/103

```

1 MEMDIDLDDIENL.....LPNTFNKYSSSSCDKKGCKTLKSGSKSPS... 42
  |:|.:.|||...|..|:| |...:| |||...:
491 IELAIKIAVKNLDEIKGHTAIFSDVSGSMSTSMGGAKKYGSVRTCLEC 540

43 .LTIPKLQKQ.....LEFYFSDANLYNDSFLRKLVLKSGEQRVEIETLL 85
  |:|.:.||| : : |...:|...:|...:| |...:|
541 ALVLGLMVKQRCCKSSFYIFSSPSSQCNCYL.EVDLPGDEL RPSMQKLL 589

```

FIG. 39

```

telomerase p43 LQKQLEFYFSDANLYNDSFLRKLVLKSGEQRVEIETLLM
human La ICHQVEYYFGDFNLPRDKFLKEQI.KLDEGWVPLEIMIK
Xenopus LaA ICEQIEYYFGDHLNLPDKFLKQOI.LLDDGWVPLETMIK
Drosophila La ILRQVEYYFGDANLNDRKFLREQIGKNEDGWVPLSVLVT
S. c. Lhplp CLKQVEFYHSEFNFPYDRFLRTTAEK.NDGWVPISTIAT

```

FIG. 41

```

1 aactcattta attactaatt taatcaacaa gattgataaa aagcagtaaa taaaacccaa
61 tagatttaatt ttagaaagta tcaattgaaa aatggaaatt gaaaacaact aagcacaata
121 gccaaaagcc gaaaaattgt ggtgggaact tgaattagag atgcaagaaa accaaaatga
181 tatataagtt aggggttaaga ttgacgatcc taagcaatat ctctgtaacg tcaactgcagc
241 atgtttgttg taggaaggta gttactacta agataaagat gaaagaagat atatcatcac
301 taaagcactt cttgaggttg ctgagtctga tcctgagttc attgtagcat tttgtgttgt
361 catccgtaat gaactttaca tcagaactac cactaactac aaagcagtac ttttgcttaa
421 ccacaagaat actcaaccat tcatcgaaaa gtacttcaac aaagcagtac ttttgcttaa
481 tgacttactg gaagtctgtg aatttgcata ggttctctat atttttgatg caactgaatt
541 caaaaatttg tatcttgata ggatactttc ataagatatt cgtaaggaac tcactttccg
601 taagtgttta caaagatgag tcagaagcaa gttttctgaa ttcaacgaat actaacttgg
661 taagtattgc actgaatcct aacgtaagaa aacaatgttc cgttacctct cagttaccaaa
721 caagtaaaaag tgggattaaa ctaagaagaa gagaaaagag aatctcttaa ccaaacttta
781 ggcaataaag gaatctgaag ataagtccaa gagagaaact ggagacataa tgaacgttga
841 agatgcaatc aaggctttta aaccagcagt tatgaagaaa atagccaaga gatagaatgc
901 catgaagaaa cacatgaagg cacctaaaat tcctaactct accttggaa caaagtactt
961 gaccttcaag gatctcatta agttctgcca tatttctgag cctaaagaaa gagtctataa
1021 gatccttggg aaaaaatacc ctaagaccga agaggaatac aaagcagcct ttggtgattc
1081 tgcacttgca cccttcaatc ctgaattggc tggaaagcgt atgaagattg aaatctctaa
1141 aacatgggaa aatgaactca gtgcaaaagg caacactgct gaggtttggg ataatttaatt
1201 ttcaagcaat taactcccat atatggccat gttacgtaac ttgtctaaca tcttaaaagc
1261 cgggtgttca gatactacac actctattgt gatcaacaag atttgtgagc ccaaggccgt
1321 tgagaactcc aagatgttcc ctcttcaatt ctttagtgcc attgaagctg ttaatgaagc
1381 agttactaag ggattcaagg ccaagaagag agaaaatatg aatcttaaag gtcaaatcga
1441 agcagtaaaag gaagtgttg aaaaaaccga tgaagagaag aaagatatgg agttggagta
1501 aaccgaagaa ggagaatttg ttaaagtcaa cgaaggaatt ggcaagcaat acattaactc
1561 cattgaactt gcaatcaaga tagcagttaa caagaattta gatgaaatca aaggacacac
1621 tgcaatcttc tctgatgttt ctggttctat gagtacctca atgtcaggtg gagccaagaa
1681 gtatgggttc gttcgtactt gtctcgagtg tgcattagtc cttgggttga tggtaaaata
1741 acgttgtgaa aagtcctcat tctacatctt cagttcacct agttctcaat gcaataagtg
1801 ttacttagaa gttgatctcc ctggagacga actccgtcct tctatgtaaa aacttttgca
1861 agagaaaagga aaacttggtg gtggtactga tttcccctat gagtgcattg atgaatggac
1921 aaagaataaa actcacgtag acaatatcgt tattttgtct gatatgatga ttgcagaagg
1981 atattcagat atcaatgtta gaggcagttc cattgttaac agcatcaaaa agtacaagga
2041 tgaagtaaat cctaacatta aaatctttgc agttgactta gaaggttacg gaaagtgcct
2101 taatctaggt gatgagttca atgaaaacaa ctacatcaag atattcggta tgagcgattc
2161 aatcttaaag ttcatttcag ccaagcaagg aggagcaaat atgggtcgaag ttatcaaaaa
2221 ctttgccctt caaaaaatag gacaaaagtg agtttcttga gattcttcta taacaaaaat
2281 ctcacccacac tttttgttt tattgcatag ccattatgaa atttaaatta ttatctattt
2341 atttaagtta cttacatagt ttatgtatcg cagtctatta gcctattcaa atgattctgc
2401 aaagaacaaa aaagattaaa a

```

FIG. 42

	Motif A	Motif B
Consensus telomerase p123 Dong (LINE) a1 S.c. (groupII) HIV-RT L8543.12	h--hDh---h--h GQPKLFFATMDIEKCYDSVNREKLSFLKTTKLL-100-KFYKQTKGIPQGLCVSSILSSFYATLEESSLGFL KNRNLHCTYDDYKKAFFDSIPHSWLIQVLEIYKIN-28-RQIAIKKGIVQGDSLSPWFCALNPLSHQLHNR FGGSNWFREVDLKKCFDTISHDLIIKELKRYISD-26-HVPVGPVRCVQGAFTSPALCNVALLRDLRRLAGLA LKKKKSVTVLVDVGDAYFSVPLDEDFRKYTAFTIP-7-GIRYQYNVLPQGWKGSPIFQSSMTKILEPFRKQN VLPELYFMKFDVKSCYDSIPRMECNRIKDALKN-68-KCYIREDGLFQSSLSAPIVDLVYDDLLLEFYSEFK	h---+-QG---sp
Consensus telomerase p123 Dong (LINE) a1 S.c. (groupII) HIV-RT L8543.12	h--YhDDh -14-LMRLTDDYLLITTTQENN-0-AVLFIKELINVSRENGFKFNMKLQT-23-QDYCDWIGISI -16-HLIYMDDIKLYAKNDKE-0-MKKLIDTTTIFSNDISMQFGLDKCKT-25-KCLYKYLGFQQ -55-YVRYADDILIGVLGSKN-2-KIIKRDNLNFLNS.LGLTINEEKTLI-4-ETPARFLGYNI -4-IVQYMDLIVGSHLEIG-1-HRTKIEELRQHLLRWGLTTPDKKHQK-0-EPPFLWMGYEL -8-ILKLADDFLIISTDQQQ.....VINIKKLAMGGFQKYNANR-41-IRSKSSKGIFR	Motif D gh-h---K h-hLgh-h

FIG. 40

51/013

MEIENNQAQQPKAEKLWWELELEMQENQNDIQVRVKIDDPKQYL
VNVTAACLLQEGSYQDKDERRYIITKALLEVAESDPEFICQLA
VYIRNELYIRTTTNYIVAFVHVHNTQPFIEKYFNKAVLLPNDL
LEVCEFAQVLYIFDATEFNLYLDRILSQDIRKELTFRKCLQRC
VRSKFSEFNEYQLGKYCTESQRKKTFRYLSVTNKQKWDQTKKK
RKENLLTKLQAIKESEDKSKRETGDIMNVEDAIKALKPAVMKKI
AKRQONAMKKHMKAPKIPNSTLESKYLTFFKDLIKFCHISEPKERV
YKILGKKYPKTEEEYKAAFSDSASAPFNPFLAGKRMKIEISKTW
ENELSAKGNTAEVWDNLISSNQLPYMAMLRNLSNILKAGVSDTT
HSIVINKICEPKAVENSKMFPLQFFSAIEAVNEAVTKGFKAKKR
ENMNLKGQIEAVKEVVEKTDEEKKDMELEQTEEGEFVKVNEGIG
KQYINSIELAIIKIAVNKNLDEIKGHTAIFSDVSGSMSTSMSSGA
KKYGSVRTCLECALVLGLMVQRCEKSSFYIFSSPSSQCNKCYL
EVDLPGEDELPRSMQKLLQEKGLGGGTDFPYECIDEWTKNKTHTV
DNIVILSDMMIAEGYSINVRGSSIVNSIKKYKDEVNPNIKIFA
VDLEGYKCLNLGDEFNENNYIKIFGMSDSLKFISAKQGGANM
VEVIKNFALQKIGQK

FIG. 43

MSRRNQKKPQAPIGNETNLDFVLQNLVYKSQIEHYKTQQQQIK
EEDLKLKFKNQDQDGNNSGNDDEENNSNKQELLRRVNIQIKQ
QVQLIKKVGSKVEKDLNLNEDENKKNGLSEQQVKEEQLRTITEE
QVKYQNLVFNMDYQLDLNESGGHRRHRETDYDEKWFIEISHDQ
KNYVSIYANQKTSYCWWLKDYFNKNNDHLNVISINRLETEAEFY
AFDDFSQTIKLTNNSYQTVNIDVNFNNLCILALLRFLSLERF
NILNIRSSYTRNQYNFEKIGELLETFVAVFSHRHLQGIHLQVP
CEAFQYLVNSSSQISVKDSLQVYSFSTDCLKLVDNKNVQDYFKF
LQEFPRLTHVSQQAIPVSATNAVENLNVLLKKVKHANLNLVSI P
TQNFDFYFVNLQHLKLEFGLEPNILTKQKLENLLLSIKQSKNL
KFLRLNFYTYVAQETSRRKQILKQATTIKNLKNNKNQEETPETKD
ETPSESTSGMKFFDHLSELTELEDFSVNLQATQEIYDSLHLLI
RSTNLKKFKLSYKEMEKSMDTFIDLKNIYETLNNLKRCVNI
SNPHGNISYELTNKDSTFYKFKLTNLQELQHAKYTFKQNEFQFN
NVKSAKIESSSLESLEDIDSLCKSIASCKNLQNVNIIASLLYPN
NIQKNPFNKPNLLFFKQFEQLKNLENVSINCILDQHIILNSISEF
LEKNKKIKAFILKRYLLQYLDYTKLFTLQQLPELNQVYINQ
QLEELTVSEVHKQVWENHKQAFYEPLCEFIKESSTQLQLIDFD
QNTVSDDSIKKILESISESKYHHYLRNPSQSSSLIKSENEEIQ
ELLKACDEKGVLVKAYYKFLCLPTGTYDYNSDRW

FIG. 45

MKILFEFIQDKLDIDLQTNSTYKENLKCCHFNGLDEILTTCFAL
PNSRKIALPCLPGDLSHKAVIDHCIYLLTGELYNVLTFGYKI
ARNEDVNNSLFCHSANVNVTLKGAAWKMFHSLVGTYAFVDLLI
NYTVIQFNGQFFTQIVGNRCNEPHLPPKWQRSSSSSATAAQIK
QLTEPVTNKQFLHKLINSSSFFPYSKILPSSSSSIKKLTDLREA
IFPTNLVKIPQRLKVRINLTQKLLKRHKRLNYSILNSICPPL
EGTVLDLSHLRQSPKERVLFKFIIVILQKLLPQEMFGSKKNKGK
IIKNLNLNLLSLPLNGYLPFDSLLKKLRLKDFRWLFISDIWFTKH
NFENLNQLAICFISWLFRLPKIIQTFFYCTEISSTVTIVYFR
HDTWNKLITPFIVEYFKTYLVENNVCRNHNSTLSNFNHSMRI
IPKKSNEEFRIIAIPCRGADEEEFTIYKENHKNAIQPTQKILEY
LRNKRPTSFTKIYSPTQIADRIKEFKQRLKKFNNVLPPELYFMK
FDVKSCYDSIPRMECMRILKDALKNENGFFVRSQYFFNTNTGVL
KLFNVVNASRVKPYELYIDNVRTVHLSNQDVINVVEMEIFKTA
LWVEDKCYIREDGLFQSSLSAPIVDLVYDDLLEFYSEFKASPS
QDTLILKLADDFLIISTDQQQVINIKKLAMGGFQKYNANANRDK
ILAVSSQSDDDTVIQFCAMHIFVKELEVWKHSSTMNHFHRSKS
SKGIFRSLIALFNTRISYKTIIDTNLNSTNTVLMQIDHVKNISE
CYKSAFKDLSINVTQNMQFHSFLQRIIEMTVSGCPITKCDPLIE
YEVRFITLNGFLESLSNNTSKFKDNIILLRKEIQHLQAYIYIYI
HIVN

FIG. 46

```

1 tcaatactat taattaataa ataaaaaaaaa gcaaactaca aagaaaatgt caaggcgtaa
61 ctaaaaaaag ccataggctc ctataggcaa tgaaacaaat cttgattttg tattacaaaa
121 tctagaagtt tacaaaagcc agattgagca ttataagacc tagtagtaat agatcaaaga
181 ggaggatctc aagcttttaa agttcaaaaa ttaagattag gatggaaact ctggcaacga
241 tgatgatgat gaagaaaaca actcaaaaaa ataataagaa ttattaagga gagtcaatta
301 gattaagtag caagtttaat tgataaaaaa agttggttct aaggtagaga aagatttgaa
361 tttgaacgaa gatgaaaaca aaaagaatgg actttctgaa tagcaagtga aagaagagta
421 attaagaacg attactgaag aatagggttaa gtattaaaat ttagtattta acatggacta
481 ccagtttagat ttaaatgaga ctgggtggcca tagaagacac agaagagaaa cagattatga
541 tactgaaaaa tggtttgaaa catctcatga ccaaaaaaat tatgtatcaa tttacgcaa
601 ctaaaagaca tcatattggt ggtggcttaa agattatttt aataaaaaaca attatgatca
661 tcttaatgta agcattaaca gactagaaac tgaagccgaa ttctatgcct ttgatgattt
721 ttcacaaaca atcaaactta ctaataattc ttactagact gttaacatag acgttaattt
781 tgataataat ctctgtatac tcgcattgct tagattttta ttatcactag aaagattcaa
841 tattttgaat ataagatctt cttatacaag aaattaatat aattttgaga aaattgggtga
901 gctacttgaa actatcttcg cagttgtctt ttctcatcgc cacttcattt
961 acaagttcct tgcgaagcgt tctaattttt agttaactcc tcatcataaa ttagcggttaa
1021 agatagctaa ttataggtat actctttctc tacagactta aaattagttg acactaacaa
1081 agtccaagat tatttttaagt tcttataaga attcctcgt ttgactcatg taagctagta
1141 ggctatccca gttagtgcta ctaacgctgt agagaacctc aatgttttac ttaaaaagggt
1201 caagcatgct aatcttaatt tagtttctat cctaccta aaattttg atttctactt
1261 tgttaattta taacatttga aattagagtt tggattagaa ccaaattttt tgacaaaaaca
1321 aaagcttgaa aatctacttt tgagtataaa ataatacaaaa aatcttaaat ttttaagatt
1381 aaacttttac acctacgttg cttagaagaa ctccagaaaa cagatattaa aacaagctac
1441 aacaatcaaa aatctcaaaa acaataaaaa tcaagaagaa actcctgaaa ctaaagatga
1501 aactccaagc gaaagcacia gtggtatgaa attttttgat catctttctg aattaacgga
1561 gcttgaagat ttcagcgtaa acttgtaagc taccgaagaa atttatgata gcttgacaaa
1621 acttttgatt agatcaacaa atttaagaa gttcaaatta agttacaaat atgaaatgga
1681 aaagagtaaa atggatacat tcatagatct taagaatatt tatgaaacct taaacaatct
1741 taaaagatgc tctgttaata tatcaaacc tcatggaaac atttcttatg aactgacaaa
1801 taaagattct actttttata aatttaagct gaccttaaac taagaattat aacacgctaa
1861 gtatactttt aagtagaacg aatttttaatt taataacggt aaaagtgcaa aaattgaatc
1921 ttcttcatta gaaagcttag aagatattga tagtctttgc aaatctattg cttcttgtaa
1981 aaattttacaa aatgttaata ttatcgccag tttgctctat cccaacaata tttagaaaaa
2041 tcctttcaat aagcccaatc ttctattttt caagcaattt gaataattga aaaatttgga
2101 aaatgtatct atcaactgta ttcttgatca gcatatactt aattctattt cagaattctt
2161 agaaaagaat aaaaaaataa aagcattcat tttgaaaaga tattatttat tacaatatta
2221 tcttgattat actaaattat ttaaaacact tcaatagtta cctgaattaa attaagttaa
2281 cattaattag caattagaag aattgactgt gagtgaagta cataagtaag tatgggaaaa
2341 ccacaagcaa aaagctttct atgaaccatt atgtgagttt atcaaagaat catcctaaac
2401 ccttttagcta atagattttg accaaaacac tgtaagtgat gactctatta aaaagatttt
2461 agaacttata tctgagtcta agtatcatca ttatttgaga ttgaacccta gttaatctag
2521 cagtttaatt aaatctgaaa acgaagaat ttaagaactt ctcaaagctt gcgacgaaaa
2581 aggtgtttta gtaaaagcat actataaatt cctctatgt ttaccaactg gtacttatta
2641 cgattacaat tcagatagat ggtgattaat taaatattag tttaaataaa tattaatat
2701 tgaatatttc tttgcttatt atttgaataa tacatacaat agtcattttt agtgttttga
2761 atatatttta gttatttaatt tcattatttt aagtaaataa ttatttttca atcatttttt
2821 aaaaaatcg

```

FIG. 44

Oxytricha
Euplotes

LCVSYILSSFYANLEENALQFLRKESMDPEKPETNLLMRLT
LCVSSILSSFYATLEESSLGFLRDESMNPENPNVLLMRLT

FIG. 47

ATTTATACTCATGAAAATCTTATTCGAGTTCATTCAAGACAAGCTTGACATTGATCTACA
GACCAACAGTACTTACAAAGAAAATTTAAAATGTGGTCACTTCAATGGCCTCGATGAAAT
TCTAACTACGTGTTTCGCACTACCAAATTCAGAAAAATAGCATTACCATGCCTTCCTGG
TGACTTAAGCCACAAAGCAGTCATTGATCACTGCATCATTTACCTGTTGACGGGCGAATT
ATACAACAACGTACTAACATTTGGCTATAAAATAGCTAGAAATGAAGATGTCAACAATAG
TCTTTTTTGGCATTCTGCAAATGTTAACGTTACGTTACTGAAAGGCGCTGCTTGAAAAAT
GTTCCACAGTTTGGTCGGTACATACGCATTTCGTTGATTTATTGATCAATTATACAGTAAT
TCAATTTAATGGGCAGTTTTCCTCAATCGTGGGTAACAGATGTAACGAACCTCATCT
GCCGCCCAAATGGGTCCAACGATCATCTCATCATCCGCAACTGCTGCGCAAATCAAACA
ACTTACAGAACCAGTGACAAATAACAATTCTTACACAAGCTCAATATAAATTCCTCTTC
TTTTTTTCTTATAGCAAGATCCTTCCTTCATCATCATCTATCAAAAAGCTAACTGACTT
GAGAGAAGCTATTTTCCACAAATTTGGTTAAAATTCCTCAGAGACTAAAGGTACGAAT
TAATTTGACGCTGCAAAAGCTATTAAAGAGACATAAGCGTTTGAATTACGTTTCTATTTT
GAATAGTATTTGCCCACCATTGGAAGGGACCGTATTGGACTTGTGCGATTGAGTAGGCA
ATCACCAAAGGAACGAGTCTTGAAATTTATCATTGTTATTTTACAGAAGTTATTACCCCA
AGAAATGTTTGGCTCAAAGAAAAATAAAGGAAAAATTATCAAGAATCTAAATCTTTTATT
AAGTTTACCCTTAAATGGCTATTTACCATTGATAGTTTGGTTGAAAAAGTTAAGATTAAA
GGATTTTCGGTGGTTGTTTCAATTTCTGATATTTGGTTCACCAAGCACAAATTTTGAAACTT
GAATCAATTGGCGATTGTTTCAATTTCTGGCTATTTAGACAATAATTCCCAAAATTAT
ACAGACTTTTTTTTACTGCACCGAAATATCTTCTACAGTGACAATTGTTTACTTTAGACA
TGATACTTGAATAAACTTATCACCCCTTTTATCGTAGAATATTTTAAGACGTACTTAGT
CGAAAACAACGTATGTAGAAACCATAATAGTTACACGTTGTCCAATTTCAATCATAGCAA
AATGAGGATTATACCAAAAAAAGTAATAATGAGTTCAGGATTATTGCCATCCCCTGACG
AGGGGCGAGCGAAGAAGAAATTCACAATTTATAAGGAGAATCACAAAAATGCTATCCGCC
CACTCAAAAAATTTTAGAATACTTAAGAAACAAAAGGCCGACTAGTTTACTAAAATATA
TTCTCAACGCAAAATAGCTGACCGTATCAAAGAATTTAAGCAGAGACTTTTAAAGAAAT
TAATAATGTCTTACCAGAGCTTTATTTTCATGAAATTTGATGTCAAATCTTGCTATGATTC
CATACCAAGGATGGAATGTATGAGGATACTCAAGGATGCGCTAAAAAATGAAAATGGGT
TTTCGTTAGATCTCAATATTTCTTCAATACCAATACAGGTGTATTGAAGTTATTTAATGT
TGTTAACGCTAGCAGAGTACCAAAACCTTATGAGCTATACATAGATAATGTGAGGACGGT
TCATTTATCAAATCAGGATGTTATAAACGTTGTAGAGATGGAAATATTTAAAACAGCTTT
GTGGGTGGAAGATAAGTGCTACATTAGAGAAGATGGTCTTTTTCAGGGCTCTAGTTTATC
TGCTCCGATCGTTGATTGTTGTTGATGACGATCTTCTGGAGTTTATAGCGAGTTTAAAGC
CAGTCCCTAGCCAGGACACATTAATTTTAAAACCTGGCTGACGATTTCCTTATAATATCAAC
AGACCAACAGCAAGTGATCAATATCAAAAAGCTTGCCATGGGCGGATTTCAAAAATATAA
TGCGAAAGCCAATAGAGACAAAATTTTAGCCGTAAGCTCCCAATCAGATGATGATACGGT
TATTCAATTTTGTGCAATGCACATATTTGTTAAAGAATTGGAAGTTTGGAAACATTCAG
CACAATGAATAATTTCCATATCCGTTCGAAATCTAGTAAAGGATATTTCGAAGTTTAAAT
AGCGCTGTTTAACTAGAAATCTCTTATAAAACAATTGACACAAATTTAAATTCACAAA
CACCGTTCTCATGCAAATGATCATGTTGTAAAGAACATTTCGGAATGTTATAAATCTGC
TATTAAGGATCTATCAATTAATGTTACGCAAAATATGCAATTTCAATTCGTTCTTACACG
CATCATTGAAATGACAGTCAGCGGTTGTCCAATTACGAAATGTGATCCTTTAATCGAGTA
TGAGGTACGATTACCATATTTGAATGGATTTTGGAAAGCCTATCTTCAACACATCAAA
ATTTAAAGATAATATCATTCTTTTGAGAAAGGAAATTCACACTTGCAAGC

FIG. 48

AKFLHWLMSVYVVVELLRSFFYVTETTFQKNRLFFYRKSVWSKLQSIGIRQHLKR
 VQLRDVSEAEVRQHREARPALLTSRLRFIPKPDGLRPVNM DYVVGARTFRREK
 RAERLTSRVKALF SVLNYERA

FIG. 49

GCCAAGTTCCTGCACTGGCTGATGAGTGTGTACGTCGTCGAGCTGCTCAGGTC
 TTTCTTTTATGTCACGGAGACCACGTTTCAAAAGAACAGGCTCTTTTCTACC
 GGAAGAGTGTCTGGAGCAAGTTGCAAAGCATTGGAATCAGACAGCACTTGAAG
 AGGGTGCAGCTGCGGGACGTGTGCGAAGCAGAGGTCAGGCAGCATCGGGAAGC
 CAGGCCCCGCCCTGCTGACGTCCAGACTCCGCTTCATCCCAAGCCTGACGGGC
 TCGCGCCGATTGTGAACATGGACTACGTGCGTGGGAGCCAGAACGTTCCGCAGA
 GAAAAGAGGGCCGAGCGTCTCACCTCGAGGGTGAAGGCACTGTTCAGCGTGCT
 CAACTACGAGCGGGCGCG

FIG. 50

MTEHHTPKSRILRFLENQYVYLCTLNDYVQLVLRGSPASSYSNICERLRSVQTSFSIFLHSTVVGF
 DSKPDEGVQFSSPKCSQSELIANVVKQMFDESFERRRNLLMKGFSMNHEDFRAMHVNGVQNDLVSTF
 PNYLISILESKNWQLLEIIGSDAMHYLLSKGSIFEALPNDNYLQISGIPLFKNVFEETVSKKRKR
 TIETSITQNK SARKEVSWNSISISRFSIFYRSSYKKFKQDLYFNLHSICDRNTVHMLQWIFPRQFG
 LINAFOVKQLHKVIPLVSQSTVVPKRLLKVYPLIEQTAKRLHRISLSKVYNHYCPYIDTHDDEKILS
 YSLKPNQVFAFLRSILVRVFPKLIWGNQRIFEIILKDLETFLKLSRYESFSLHYLSNLIKISEIEWL
 VLGKRSNAKMCLSDFEKRKQIFAEFIYWLNSFIIPILOSFYITESSDLNRNTVYFRKDIWKLLCR
 PFITSMKMEAFEKINENNVRMDTQKTTLPPAVIRLLPKKNTFRLITNLKRFLIKMGSNKKMLVSTN
 QTLRPVASILKHLINEESSGIPFNLEVYMKLLTFKKDLLKHRMFGRRKYFVRIDIKSCYDRIKQDLM
 FRIVKKKLDPEFVIRKYATIHATSDRATKNFVSEAFSYFDMVPFEKVQLLSMKTSDTLFVDFVDY
 WTKSSSEIFKMLKEHLSGHIVKIGNSQYLQKVGIPQGSILSSFLCHFYMEDLIDEYLSFTKKKGSVL
 LRVVDDFLFITVNKKDAKKFLNLSLRGFEKHNFTSLEKTVINFENSNGIINNTFFNESKKRMPFFG
 FSVNMRS LDTLLACPKIDEALFNSTSVELTKHMGKSFFYKILRSSLASFAQVFIDITHNSKFNSCCN
 IYRLGYSMCMRAQAYLKRMKDIFIQRMFITDLNVIGRKIWKLAELGYTSRRFLSSAEVKWLFC
 LGMRDGLKPSFKYHPCFEQLIYQFQSLTDLIKPLRPVLRQVLF LHRRIAD

FIG. 51

ggtaccgatttacttcttcttcataagctaaattgcttccclcgaaacgctcctaaatctctggaataatattttacaaga
 actcaataacaataaccaagtcacaattcccaatgatgaagtggtatttagtgatcgataaataatttctatttttatcggtcgtta
 ccaagtataaggacaaaagaacaacttcttcccccaagacttttactttattatttccaaatatatttctcg
 ggttcgcttacttttaactcggtggtactgtttacaaactccttcgatgagactatattagattcattacagtcogtgcataatc
 agctcttgaggtagctcacagaaaactccttcaaaactccttcgatgagactatattagattcattacagtcogtgcataatc
 ttaacatggagccttacactttagatgagtcacgtcgcatgagatatttggtatcatcccaacgtttgcttggaaaag
 gttgataaattatttgcaaaatcatgtcccttagtggtgtaalcctcgaaagtlttttgatgcttgccacacgtctagcatg
 attgagataattcaaaaatttctaccactacaactccttttaacgctgtttattttctatttctattctcattggtt
 ccaaatatgtatcatctcgtattaggcttttttccgttttttactcctggaaatcgtaaccttttcaactattccccctaataga
 ataacttaaataggtttcgcttataatgtatagtagtaglagaagaatctggtgattctactcgtgaatgttatttagtttaa
 gatactttgcaaaacatttattagctatcatattataaaaaaaatcctataataataataataataatatttgcggtc
 actatattttaaaacgttatgatcagtaggacactttgcatatatattagttatgcttaagtgttaacttgaacttgcat
 GACCGAACCCATACCCCAAAAGCAGGATTCTTCGCTTTCTAGAGAAACAATATGATATACCTATGTACCTTAAATGATT
 ATGTACAACCTTGTGAGAGGGTCGCCGGAAGCTCGTATAGCAATATATGCGAACGCTTGAGAACGGATGTACAAAG
 TCCTTTTCTATTTTCTTCATTCGACTGTAGTCGGCTTCGACAGTAAGCCAGATGAAGGTGTTCAATTTTCTTCCAAA
 ATGCTCACAGTCAGAGgtatatatttttggtttggatttttctatcgggatatagtaataatgggcagCTAATAGC
 GAATGTTGTAACACAGATGTTTCGATGAAAGTTTGGCGTCGAAGGAATCTACTGATGAAGGGTTTCCATGgttaagg
 attctaattgtgaaatatttacctgcaattactgtttcaagagattgtatttaaccgataaagaATCATGAAGATTTC
 GAGCCATGCATGTAAACGAGTACAAAATGATCTCGTTTCTTAATACCTTATATCTATCTACTTGAGTCAAAA
 AATTGGCAACTTTTGTAGAAATgtaaataccggttaagattgtgcgcactttgaacaagactgacaagtatagtATCGG
 CAGTGTGCCATGCCATTACTTACTTATATCCAAAGGAAGTATTTTGGGCTCTTCCAAATGACAAATACCTTCAGATTTCTG
 GCATACCCACTTTTAAAAATAAATGTGTTTGGGAACTGTGTCAAAAAAAGAAAGCAACCATTGAACACATCCATTACT
 CAAAATAAAGCGCCCGCAAGAGTTTCTGGAATAGCATTTCAATTAGTAGGTTTAGCATTTTTCACAGGTCACTCTATTT
 TAAGAAGTTTAAGCAAGTaaactaatactgtatccttcaactaattttagATCTATATTTTAACTTACACTCTATTT
 GTGATCGGAACACAGTACATGTGGCTTCATGGAATTTTCCAAAGCAATTTGGACTTATAAACGCATTTCAAGTGAAG
 CAATTGCACAAAGTATCCACTGGTATCACAGAGTACAGTTGTGCCCAACGCTCTCTAAAGGTATACCCTTTAAATTGA
 ACAACAGCAAAAGCGACTCCATCGTATTTCTCTATCAAAAGTTTACACCATTTATGCCCATATATTGACACCCACGATG
 ATGAAAAAATCCTTAGTTATTCCTTAAAGCCGAACAGGTGTTTGGGTTTCTTCGATCCATTCCTTGTTCGAGTGTTCCT
 AAATTAATCTGGGTAAACCAAGGATATTGAGATAATATAAAGgtatgtataaaatttaccactaacgatttt
 accagACCTCGAAACTTTCTTGAAATTAICGAGATACGAGTCTTTTAGTTTACATTATTAATGAGTAACATAAAGgtaa
 tatgcaaaattttttaccatttaataacaatcagATTTCAGAAATGAAATGGCTAGTCTCTGGAAAAAAGGTCAATGCG
 AAAATGTGCTTAAGTGATTTTGGAAACCGCAAGCAAAATATTTCGGGAATTCATCTACTGCTGTAACAATTCGTTTATAAT
 ACCTATTTTACAACTCTTTTATATCACTGAATCAAGTGATTTACGAAATCGAAATCTGCTGCTATACAAATTCGTTTATAAT
 GGAAACTCTGTGCGGACCCCTTTATACATCAATGAANAATGGAAGCGTTTGAAAAAATAAAGAGgtattttaagattt
 ttttgcaaaaagctaatattttcagAACAAATGTTAGGATGGATCTCAGAAAAACTACTTTGCTCCAGCAGTATTTCGTC
 TATTACCTAAGAAATACCTTTTCGTCCTCATTAAGAAATTAAGAAAAAGATTCTTAATAAAGgtatttaatttttgggtcat
 caatgtactttacttctaactattattagcagATGGTTCAACAAAAAATGTTAGTCAGTACGAACCAAACTTTACG
 ACCGTGGCATCGATACGTAAACATTTTAATCAATGAAGAAAGTAGTGGTATTTCCATTTTAACTTGGAGGTTTACATGNAAGC

FIG. 52

TTCTTACTTTTAAAGAGGATCTTCTTAAGCACCGAATGTTTGGTgaattatataaagtgcgcatctccattattatttt
 gcagcggtaagaaagtatttGTACGGATAGATATAAATCCTGTATTATGATCGAATAAAGCAAGATTTGATGTTTCGGATT
 GTTAAAAAGAAACCTCAAGATCCCGAATTTGTAATTCGAAAGTATGCAACCATACATGCAACAGTGACCGAGCTACAAA
 AAACCTTTGTAGTAGGCGTTTTCCTATTGtaagttatttttccatttggaatttttaacaaattcttttttagTTGAT
 ATGGTCCCTTTTGAANAAGTCGTGCAGTACTTTTCTTATGAAAAACATCAGATACCTTTGTTTGTGATTGTTGGAATTATTG
 GACCAAAAGTCTTCTCGAATTTTAAATGCTCAAGGAACATCTCTTGACACATTTGTTAAAGgtataccaatttgtga
 atgtataacactaatgaacactagATAGGAAATTTCTCAATACCTTCAANAAGTTGGTATCCCTCAGGGCTCAATTTCTGT
 CATCTTTTGTGTCTATTCTATATGGAAGATTGATTGATGAATACCTATCTGTTTACGAAAAAGAAAGGATCAGTGTG
 TTACGAGTAGTCGACGATTCTCTTTTATACAGTTAATAAANAAGATGCAAAAAAATTTTGAATTTTATCTTTAAGAGG
 tgaagtgcgtcatcttaagttctaacggtgaagGATTTGAGAAAAACACAAATTTTCTACGAGCCTGGAGAAAAACAGTA
 ATAACTTTGAAAAATAGTAATGGGATAATAAACAAATACCTTTTAAATGAAAGCAAGAAAGAAATGCCATTTCTTCGGTTT
 CTCTGTGAACATAGAGTCTTTGATACATTTGTAGCATGCTCCATAAATTTGATGAAGCCTTATTTAACTCTACATCTGTAG
 AGCTGACGAAACATATGGGAAATCTTTTACAAAATTTCTAAGgtatactgtgtaactgaataaagctgacaaata
 atcagATCGAGCCTTGCAATCTTGCACAAAGTATTATTGACATTACCCACAAATTTCAAAATTTCAATTTCTTGGTGCATAT
 ATATAGGCTAGGATACCTATGTGTATGAGAGCACAGCATACTTAAAAAGGATGAAGGATATATTTTATTTCCCAAGAA
 TGTTCATAACGGgtgagtacttatttaactagaagaagcttaataaaccttagATCTTTTGAATGTTATTGGAAGAAA
 AATTTGGAAAAAGTTGGCGAAATATTAGGATATACGAGTAGGGTTCCTTGTCTCTGCAGAAAGTCAATGgtacgtgt
 cggctcgcagacttcagcaatatgtacacatcagGCTTTTGTGCTTGGAAATGAGAGATGGTTGAAACCCCTTTTCAAA
 TATCATCCATGCTTCGAACAGCTAATATACCAATTTTCAGTCAATTTGACTGATCTTATCAAGCCGCTAAGACCAGTTTTCG
 ACAGGTGTTATTTTACATAGAAGATAGCTGATTAATgtcattttcaatttatataacatccctttactggtgtc
 ttaacaatatattactaagtatagctgaccccaagcaagcatactataggatttctagtaaaagtaaaatataatctc
 gtattagttttgattgactgtctttatccttatacttttaagaagaagttgacagtggttgcactactgcccacatg
 cccattaaacgggaggtggttaaacattaaaagtaacatgaggttaactgtccttcccttcatattagaataaaggaaagtgtt
 tctataatgaataatgccgcactaatgcaaaaagacgaagattatcttcaacaagggtgattaaagcatatccgaagg
 aaaaagagtaataataccccagtggttgaagaagaaggaataatttggaacaagcttctgcagatgacaggtctaaatt
 ttggtgaccgaatttttgtaaaagccccaggttatccatggtggccgacctgtctactgagacgaaaaaactaaaggat
 agtttgaaataactaatagctcatattatgtcttatataagggtttgttttttctgacttcaattttgcatgggtgaaaag
 aaataggttaagccattatggattccgaaatagccaaatttcttggttccctcaaaagcgggaagtctaaagaacttatg
 aagcttatagggttcaaaaactcctcctgatttaaaaggaggaatcttccaccgatgaggaaatggtagcttatcagct
 gctgaggagaagcctaatttttgcaaaaagaaataatcatggggagacatctcttgatgaatcagatgcggagagat
 ctccagcggatccttgatgtcaataacttctatttctgaaatgtatgggtcctactgtcgtctcgtacttctcgtagctcta
 cgcagtttaagtgaaccaaaggtaacc

FIG. 52
(CONTINUED)

EST2 pep	FFYCTEISST VTIVYFRHDT WN-----KLIT P-----FIVE YFK-TYLVEN	40
Euplotes pep	FFVTEQOKS YSKTYRKN IWDVI-MKMS IAD-----LKK ETLA--EVQE	43
Trans of tetrahymen	-----KHKE GSQIFYRKP IWKLVSCLTI VKVRIQFSEK NKQMKNNFYQ	44
Consensus	FFY.TE..K. .S..YYRK. IW....-KL..-F..KV..	50
EST2 pep	NVCRNHSY- ----- TLSNFNHSM RIIPKKSNNNE FRITAIPCRG	79
Euplotes pep	KEVEEWKSL ----- ---GFAPQKG RIIPKKT-- FRPIMTFNKK	78
Trans of tetrahymen	KIQLEENLE KVEEKLIPED SFQYPQGL RIIPKKS-- FRPIMTFLRK	92
Consensus	K...E..... -----F..GKL RIIPKK...-- FRPIMTF.RK	100
EST2 pep	ADEEFTIYK ENHKNAIQPT QKILEYLRNK RPTSFTKIYS PTOIADRIKE	129
Euplotes pep	IVNSDRKTTK LTTNTKLLNS HMLKTLKN- -----RMFK -DPFGFAVN	120
Trans of tetrahymen	DKQNIK--- LNLNQILMDS QLVFRNLKD- -----ML-G -QKIGYSVFD	130
ConsensusK..K LN.N..L..S QL.L..LKN- -----.... -..IG..VF.	150
EST2 pep	FKQRLLRNF NVL----- -PELYFMKFD VKSCYD	157
Euplotes pep	YD-DVMKKYE EFVCKWKQVH CPKLFATMD IEKCYD	155
Trans of tetrahymen	NK-QISEKFA QFIEKWKNG RPLYYVTL- -----	158
Consensus	.K-....KKF. .F..KWK..G .B.LYF.T.D ...CYD	186

FIG. 53

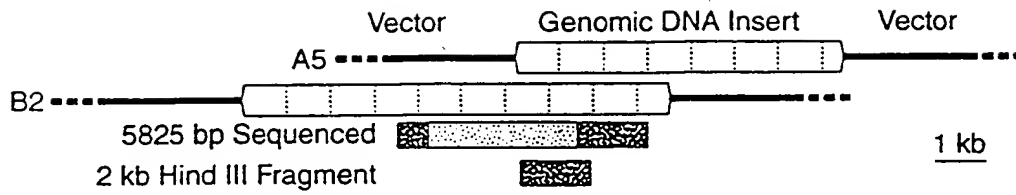


FIG. 55A

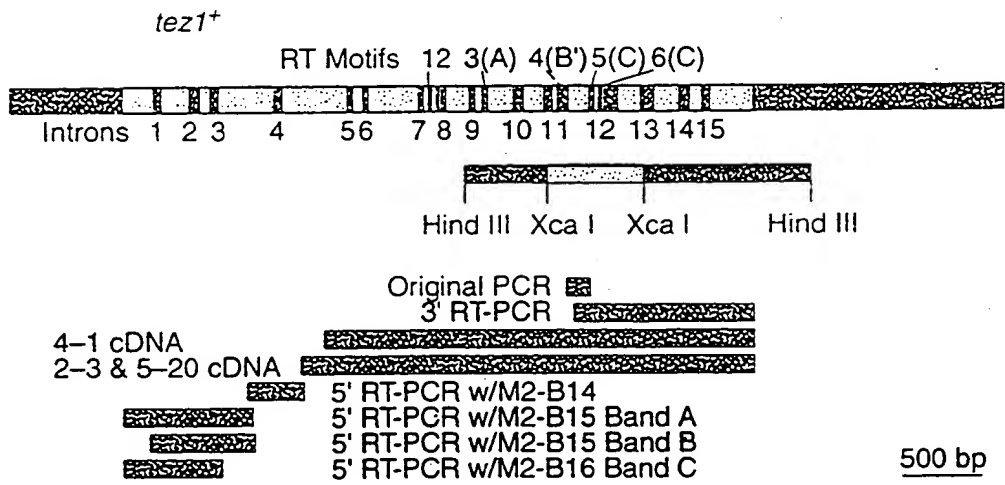


FIG. 55B

59/103

S-1: FFY VTE TTF QKN RLF FYR KSV WSK
S-2: RQH LKR VQL RDV SEA EVR QHR EA
S-3: ART FRR EKR AER LTS RVK ALF SVL NYE

A-1: AKF LHW LMS VYV VEL LRS FFY VTE TTF Q
A-2: LFF YRK SVW SKL QSI GIR QHL KRV QLR DVS
A-3: PAL LTS RLR FIP KPD GLR PIV NMD YVV

FIG. 54

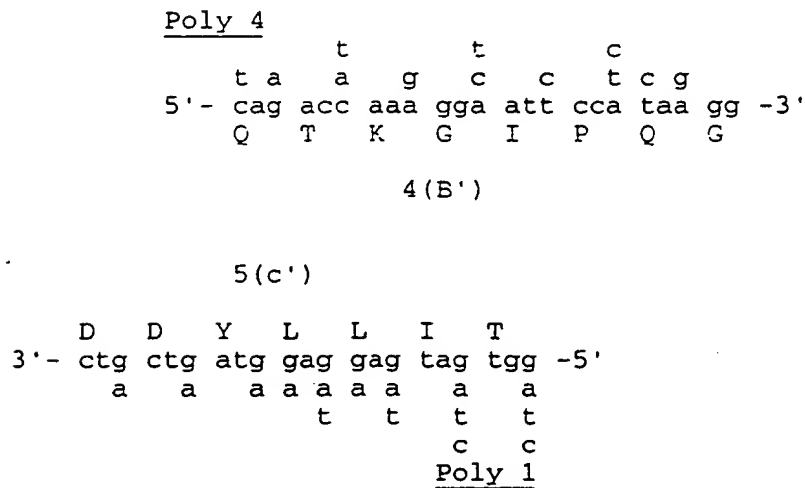


FIG. 56

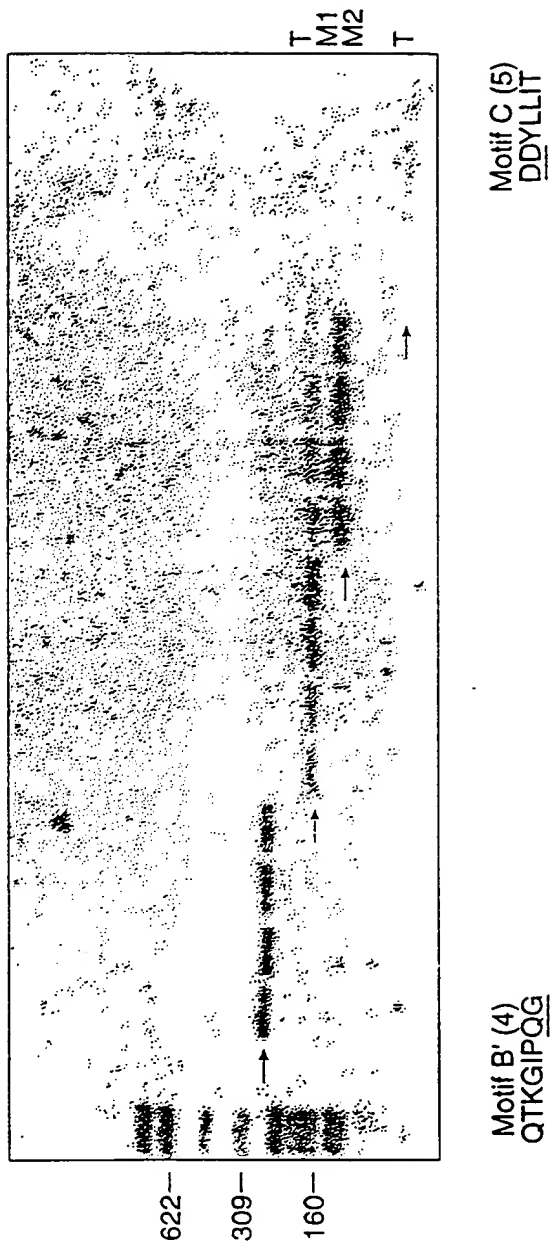


FIG. 57

Ot	LCVSYILSSFYYANLEENALQFLRKESMDPEKPTNLLMRLT	
Ea_p123	KGIPOGLCVSSILSSFYYATLEESSLGFLRDESMNPENPNVNLMLRLTDDYLLIT	
Sp_M2	SILSSFCHFYMEDLIDEYLSFTKKK-----GSVLLRVV	
Sc_p103	DGLFQGSLSAPIVDLVYDDLLLEFYSEFKASPS-----QDTLILKLADDFLIIS	
	* *	
		<----Actual Genomic Sequence.

Q	K	V	G	I	P	Q	G
caa	aaa	ggt	ggt	atc	cct	cag	gg.....

Poly 4

t	t	c
t	a	a
g	c	c
c	t	c
g	g	g
caa	acc	aaa
gga	att	cca
taa	gg	----

ag	acc	aaa	gga	att	cca	tca	ggc	tca	att	ctg	tca	tct	ttt	ttg	tgt	cat	ttc	tat	atg
tc	tgg	ttt	cct	taa	ggt	agt	ccg	agt	taa	gac	agt	aga	aaa	aac	aca	gta	aag	ata	tac

K	G	I	P	S	G	S	I	L	S	S	F	L	C	H	F	Y	M
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

FIG. 58

GAA GAT TTG ATT GAT GAA TAC CTA TCG TTT ACG AAA AAG AAA GGA TCA GTG TTG TTA CGA
 CTT CTA AAC TAA CTA CTT ATG GAT AGC AAA TGC TTT TTC TTT CCT AGT CAC AAC AAT GCT
 E D L I D E Y L S F T K K G S V L L R

GTA GTC gac gac tac ctc ctc atc acc
 CAT CAG ctg ctg atg gag gag tag tgg

V V D D Y L L I T

<----- ctg ctg atg gag gag tag tgg
 a a a a a a a
 t t t t t
 C C
Poly 1

.....gac gat ttc ctc ttt ata aca..... <----Actual Genomic Sequence
 D D F L F I T

FIG. 58
 (CONTINUED)

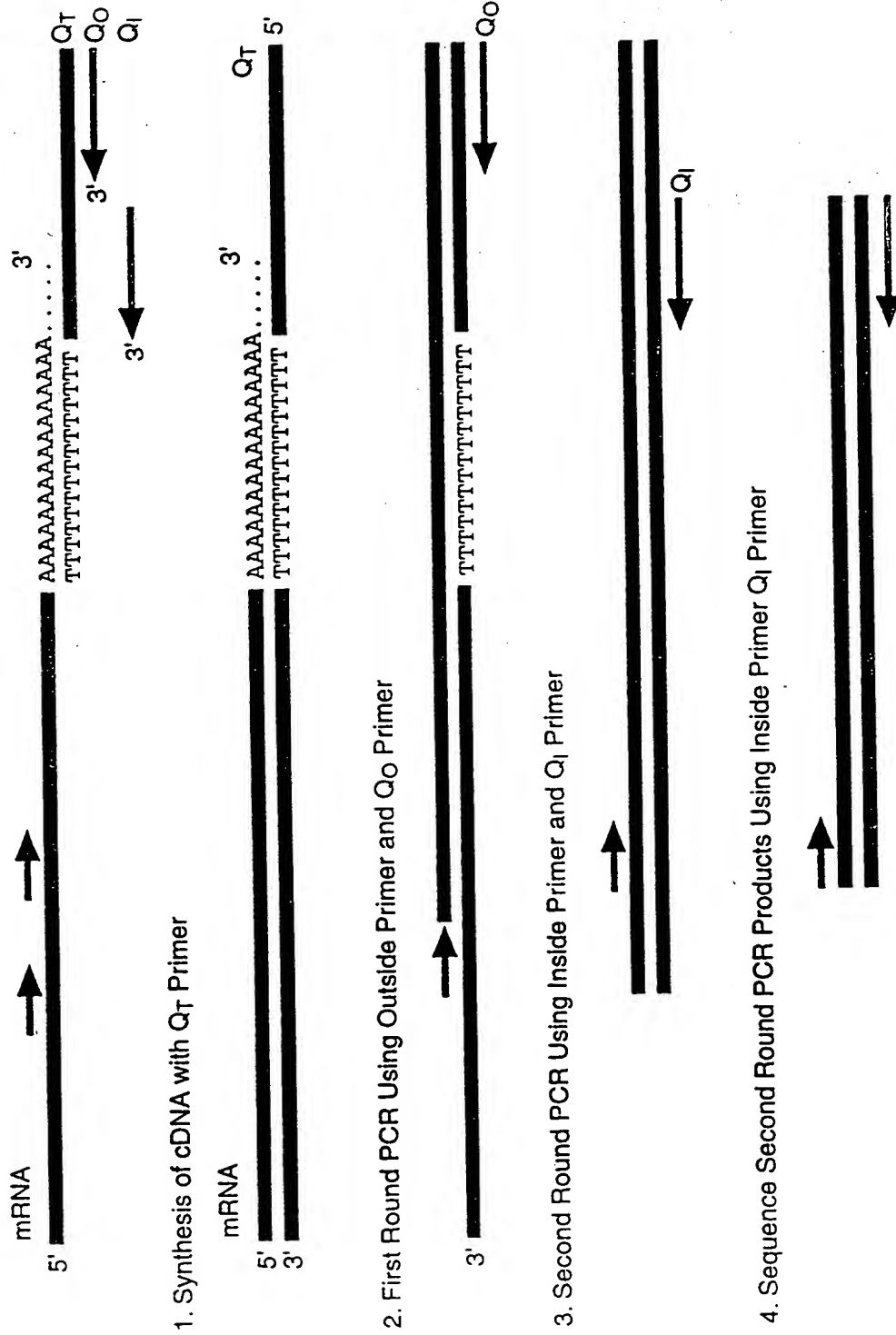


FIG. 59

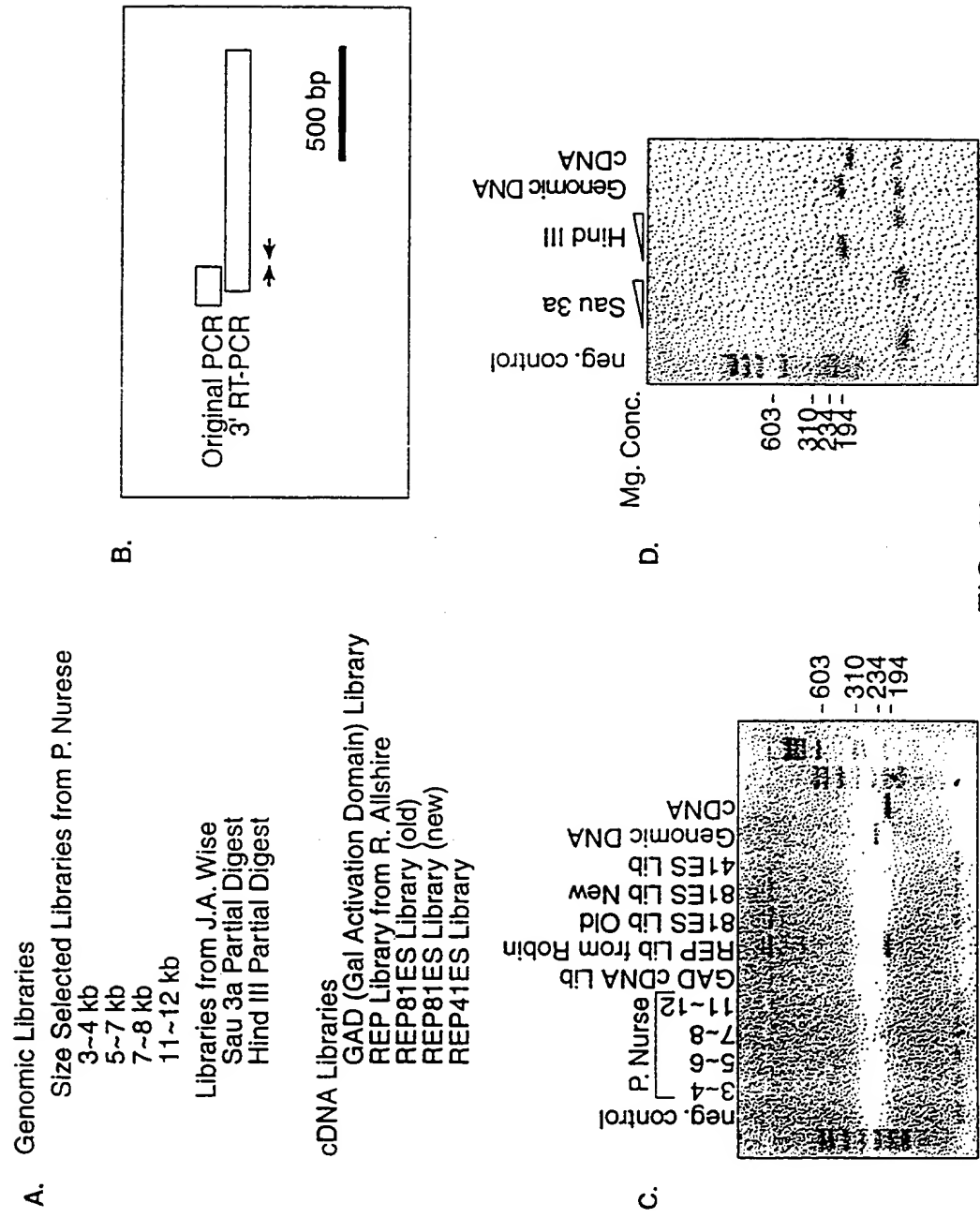


FIG. 60

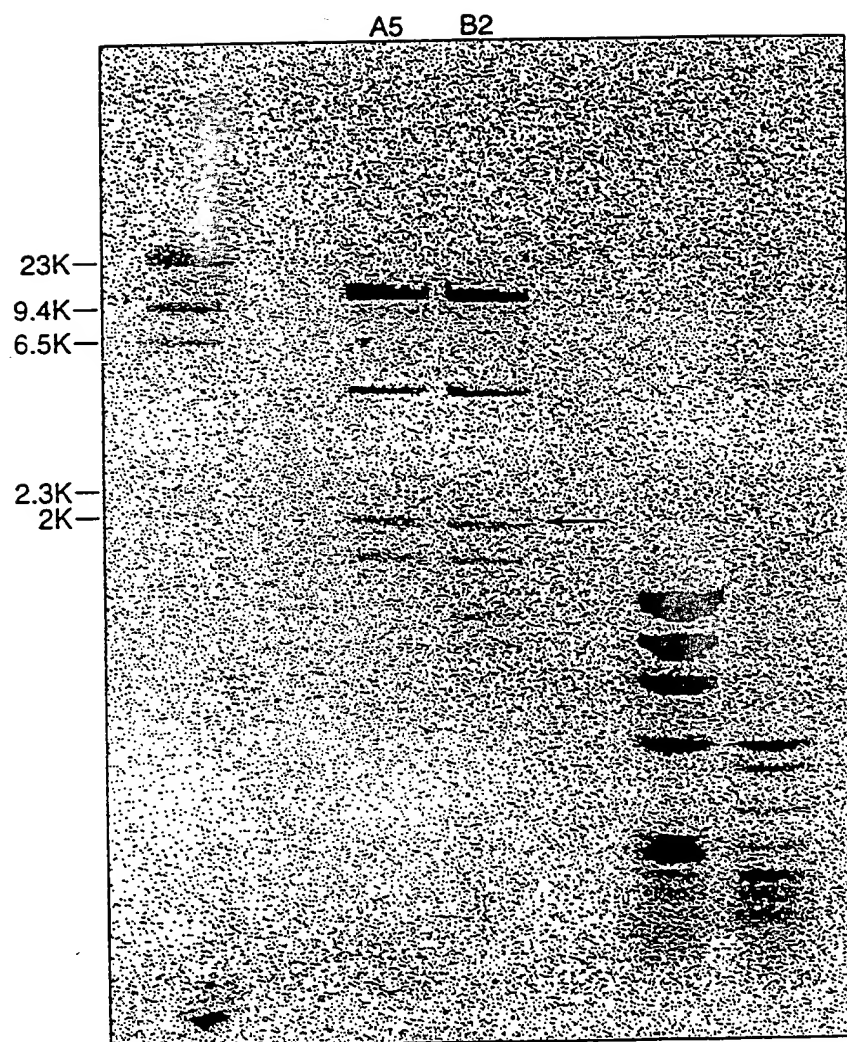


FIG. 61

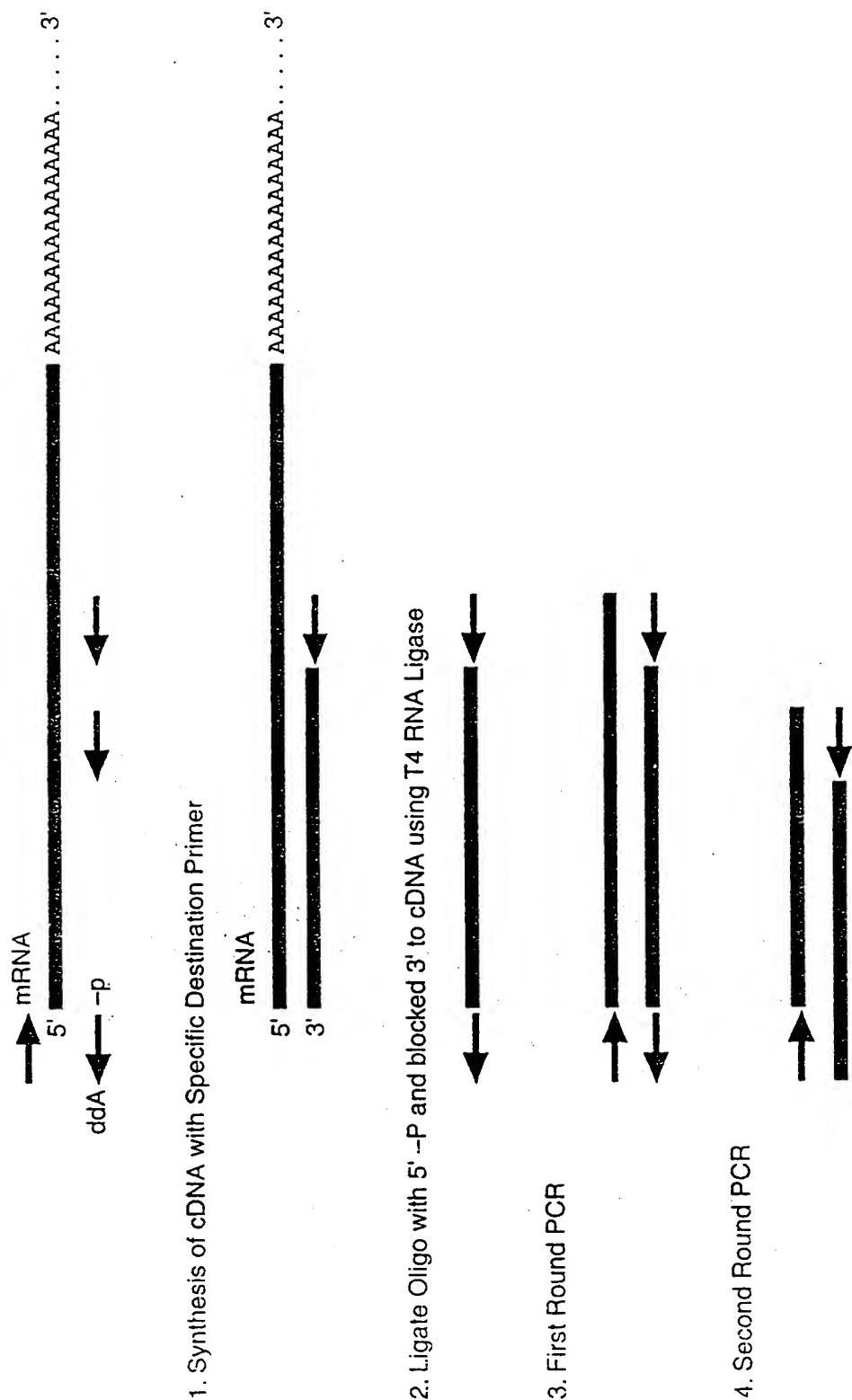


FIG. 62

Motif O

S.p. Tez1p	(429) .	WLYNSFIIPILQSFYITESSDLNRNRTVYFRKDIW	... (35) ...
S.c. Est2p	(366) .	WLFRLIPKIIQTFYCTEISSTVT-IVYFRHDTW	... (35) ...
E.a. p123	(441) .	WIFEDLVVSLIRCFYVTEQQKSYSKTYYYRKNIW	... (35) ...
		*** ** *	
		Motif 1	Motif 2
		p hh h k	hR h
S.p. Tez1p		AVIRLLPKK--NTFRLITN-LRKRF	... (61) ...
S.c. Est2p		SKMRIIPKKSNNFRIIAIPCRGAD	... (62) ...
E.a. p123		GKLRLLPKK--TTFRPIMTFNKKIV	... (61) ...
		* *** ** *	
		Motif 3(A) AF	
		h hDh GY h	
S.p. Tez1p		KKYFVRIDIKSCYDRIKQDLMFRIVK	... (89) ...
S.c. Est2p		ELYFMKFDVKSCYDSIPRMECMRILK	... (75) ...
E.a. p123		KLFFATMDIEKCYDSVNREKLSTFLK	... (107) ...
		* *** ** *	
		Motif 4(B')	
		hPQG pp hh h	
S.p. Tez1p		YLQKVGIPQGSILSSFLCHFYMEDLIDEYLSF	... (6) ...
S.c. Est2p		YIREGLFQGSLSAPIVDLVYDDLLEFYSEF	... (8) ...
E.a. p123		YKQTKGIPQGLCVSSILSSFYATLEESSLGF	... (14) ...
		* *** ** *	
		Y Motif 5(C)	Motif 6(D)
		h F DDhhh	Gh h cK h
S.p. Tez1p		VLLRVVDDFLFITVNKKDAKFLNLSLRGFEKHNFTSLEKTVINFENS	... (205)
S.c. Est2p		LILKLADDFLIISTDQQQVINIKKLAMGGFQKYNANAKNRDKILAVSSQS	... (173)
E.a. p123		LLMRLTDDYLLITTTQENNAVLFIKLINVSRENGFKFNMKKLQTSFPLS	... (209)
		** * *	*

FIG. 63

A.

[illegible]

FIG. 64

A.	Sp_Tip1p	219	WNSISISRFSIFYS	SS	Y	K	K	F	K	Q	D	L	Y	E	N	L	H	S	I	C	D	251																
	Sc_Est2p	184	N	-	-	-	-	-	-	-	-	K	Q	F	L	H	K	L	N	I	N	S	S	F	F	P	200											
	Ea_p123	218	N	E	K	-	-	D	H	F	L	N	N	I	N	V	P	N	W	N	N	M	K	S	R	T	R	I	F	Y	C	T	H	E	N	248		
	Sp_Tip1p	252	R	N	T	V	H	M	W	L	Q	W	I	F	P	R	Q	F	G	L	I	N	A	F	Q	V	K	Q	L	H	K	V	I	P	L	284		
	Sc_Est2p	201	-	-	-	-	-	-	-	-	-	Y	S	K	I	L	P	S	S	-	-	-	S	I	K	K	L	T	D	L	R	E	A	I	F	P	223	
	Ea_p123	249	R	-	-	-	-	-	-	-	-	N	N	Q	F	F	E	K	H	E	F	V	S	N	K	N	N	I	S	A	M	D	R	A	Q	T	I	275
	Sp_Tip1p	285	V	S	-	-	-	Q	S	T	V	V	P	K	R	L	K	V	Y	P	L	I	E	Q	T	A	K	R	L	H	R	I	S	313				
	Sc_Est2p	224	T	N	-	-	-	L	V	K	I	P	Q	R	L	K	V	R	I	N	L	T	L	Q	K	L	L	K	R	H	K	R	L	N	252			
	Ea_p123	276	F	T	N	I	F	R	F	N	R	I	R	K	K	L	K	D	K	V	I	E	K	I	A	Y	M	L	E	K	V	K	D	F	N	308		
	Sp_Tip1p	314	L	S	K	V	Y	N	H	Y	C	P	Y	I	D	-	T	H	D	D	E	K	I	L	S	Y	S	L	K	P	N	Q	-	-	-	342		
	Sc_Est2p	253	Y	V	S	I	L	N	S	I	C	P	P	L	E	G	T	V	L	D	L	S	H	L	S	R	Q	S	P	K	E	R	-	-	-	282		
	Ea_p123	309	F	N	Y	Y	L	T	K	S	C	P	L	P	E	N	W	R	E	R	K	Q	K	I	E	N	L	I	N	K	T	R	E	E	K	341		
	Sp_Tip1p	343	-	-	-	-	-	-	-	-	-	-	-	-	-	-	V	F	A	F	L	R	S	I	L	V	R	V	E	P	K	L	I	359				
	Sc_Est2p	283	-	-	-	-	-	-	-	-	-	-	-	-	-	-	V	L	K	F	I	I	V	L	L	Q	K	L	L	P	Q	E	M	299				
	Ea_p123	342	S	K	Y	Y	E	E	L	F	S	Y	T	T	D	N	K	C	V	T	Q	F	I	N	E	F	F	Y	N	I	L	P	K	D	F	374		
	Sp_Tip1p	360	W	G	N	Q	R	I	F	E	I	I	L	K	D	L	E	T	F	L	K	L	S	R	Y	E	S	F	S	L	H	Y	L	M	S	392		
	Sc_Est2p	300	F	G	S	K	K	N	K	G	K	I	I	K	N	L	N	L	L	S	L	P	L	N	G	Y	L	P	F	D	S	L	L	K	332			
	Ea_p123	375	L	T	G	-	R	N	R	K	N	F	Q	K	K	V	K	Y	V	E	L	N	K	H	E	L	I	H	K	N	L	L	L	E	406			
	Sp_Tip1p	393	N	I	K	I	S	E	I	E	W	L	V	L	G	K	R	S	N	A	K	M	C	L	S	D	F	E	K	R	K	I	F	A	425			
	Sc_Est2p	333	K	L	R	L	K	D	F	R	W	L	F	I	S	-	-	D	I	W	F	T	K	H	N	F	E	N	L	N	Q	L	A	I	362			
	Ea_p123	407	K	I	N	T	R	E	I	S	W	M	Q	V	E	T	S	-	A	K	H	E	Y	Y	F	D	H	E	N	-	I	Y	V	L	W	437		

FIG. 64
(CONTINUED)

A.	Sp_Tip1p	426	E	F	I	Y	W	L	Y	N	S	F	I	I	P	I	L	Q	S	F	F	Y	I	T	E	S	S	D	L	R	N	R	I	T	V	Y	458
	Sc_Est2p	363	C	F	I	S	W	L	F	R	Q	L	I	P	K	I	I	Q	T	F	F	Y	C	T	E	I	S	S	T	V	T	-	I	V	Y	394	
	Ea_p123	438	K	L	L	R	W	I	F	E	D	L	V	V	S	L	I	R	C	F	F	Y	V	T	E	Q	Q	K	S	Y	S	K	T	Y	Y	470	
	Sp_Tip1p	459	F	R	K	D	I	W	K	L	L	C	R	P	F	I	T	S	M	K	M	E	A	F	E	K	I	N	E	N	N	V	R	M	D	491	
	Sc_Est2p	395	F	R	H	D	T	W	N	K	L	I	T	P	F	I	V	E	Y	F	K	T	Y	L	V	E	N	N	V	C	R	N	H	N	S	427	
	Ea_p123	471	Y	R	K	N	I	W	D	V	I	M	K	M	S	I	A	D	L	K	K	E	T	L	A	E	V	Q	E	K	E	V	E	E	W	503	
	Sp_Tip1p	492	T	Q	K	T	T	L	P	A	V	I	R	L	L	P	K	K	-	-	N	T	F	R	L	I	T	N	L	R	K	R	F	L	522		
	Sc_Est2p	428	Y	T	L	S	N	F	N	H	S	K	M	R	I	I	P	K	K	S	N	N	E	F	R	I	I	A	I	P	C	R	G	A	D	460	
	Ea_p123	504	K	K	S	L	G	F	A	P	G	K	L	R	L	I	P	K	K	-	-	T	T	F	R	P	I	M	T	F	N	K	K	I	V	534	
	Sp_Tip1p	523	I	K	M	G	S	N	K	K	M	L	V	S	T	N	Q	T	L	R	P	V	A	S	I	L	K	H	L	I	N	E	-	-	-	552	
	Sc_Est2p	461	E	E	E	-	-	F	T	I	Y	K	E	N	H	K	N	A	I	Q	P	T	Q	K	I	L	E	Y	L	R	N	K	R	P	T	491	
	Ea_p123	535	N	S	D	-	-	R	K	T	T	K	L	T	T	N	T	K	L	N	S	H	L	M	L	K	T	L	K	N	R	-	M	F	564		
	Sp_Tip1p	553	E	S	S	G	I	P	F	N	L	E	V	Y	M	K	L	L	T	F	K	K	D	L	L	K	H	R	M	F	G	R	-	K	K	584	
	Sc_Est2p	492	S	F	T	K	I	Y	S	P	T	Q	I	A	D	R	I	K	E	F	K	Q	R	L	L	K	K	F	N	N	V	L	P	E	L	524	
	Ea_p123	565	K	D	P	F	G	F	A	V	F	N	Y	D	D	V	M	K	K	Y	E	E	F	V	C	K	W	K	Q	V	G	P	K	L	597		
	Sp_Tip1p	585	Y	F	V	R	I	D	I	K	S	C	Y	D	R	I	I	K	Q	D	L	M	F	R	I	V	K	K	K	L	K	D	P	E	-	F	616
	Sc_Est2p	525	Y	F	M	K	F	D	V	K	S	C	Y	D	S	I	P	R	M	E	C	M	R	I	L	K	D	A	L	K	N	E	N	G	F	557	
	Ea_p123	598	F	F	A	T	M	D	I	E	K	C	Y	D	S	V	N	R	E	K	L	S	T	F	L	K	T	T	K	L	L	S	S	D	F	630	
	Sp_Tip1p	617	V	I	R	K	Y	A	T	I	H	A	T	S	D	R	A	T	K	N	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	634	
	Sc_Est2p	558	F	V	R	S	Q	Y	F	F	N	T	N	T	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	570	
	Ea_p123	631	W	I	M	T	A	Q	I	L	K	R	K	N	N	I	V	I	D	S	K	N	F	R	K	K	E	M	K	D	Y	F	R	Q	K	663	

FIG. 64
(CONTINUED)

A.	Sp_Tip1p	635	F	V	S	E	A	F	S	Y	F	D	M	V	P	F	E	K	V	V	Q	L	L	S	-	-	M	K	T	S	D	T	L	F	V	665		
	Sc_Est2p	571	-	-	-	-	-	-	-	-	-	-	-	V	L	K	L	F	N	V	V	N	A	S	R	-	-	V	P	K	P	Y	E	L	Y	I	591	
	Ea_p123	664	F	Q	K	I	A	L	E	G	G	Q	Y	P	T	L	F	S	V	L	E	N	E	Q	N	D	L	N	A	K	K	T	L	I	V	696		
	Sp_Tip1p	666	D	F	V	D	Y	W	T	K	S	S	S	E	I	F	K	M	L	K	E	H	L	S	G	H	I	V	K	I	G	N	S	Q	Y	698		
	Sc_Est2p	592	D	N	V	R	T	V	H	L	S	N	Q	D	V	I	N	V	V	E	M	E	I	F	K	T	A	L	W	V	E	D	K	C	Y	624		
	Ea_p123	697	E	A	K	Q	R	N	Y	F	K	K	D	N	L	L	Q	P	V	I	N	I	C	Q	Y	N	Y	I	N	F	N	G	K	F	Y	729		
	Sp_Tip1p	699	L	Q	K	V	G	I	P	Q	G	S	I	L	S	S	F	L	C	H	F	Y	M	E	D	L	I	D	E	Y	L	S	F	T	K	731		
	Sc_Est2p	625	I	R	E	D	G	L	F	Q	G	S	S	L	S	A	P	I	V	D	L	V	Y	D	D	L	L	E	F	Y	S	E	F	K	A	657		
	Ea_p123	730	K	Q	T	K	G	I	P	Q	G	L	C	V	S	S	I	L	S	S	F	Y	A	T	L	E	E	S	S	L	G	F	L	R	762			
	Sp_Tip1p	732	K	K	G	-	-	-	-	-	-	-	-	-	-	-	S	V	L	L	R	V	V	D	D	F	L	F	I	T	V	N	K	K	D	A	K	756
	Sc_Est2p	658	S	P	S	Q	D	-	-	-	-	-	-	-	-	-	T	L	I	L	K	L	A	D	D	F	L	I	I	S	T	D	Q	Q	V	I	N	684
	Ea_p123	763	D	E	S	M	N	P	E	N	P	N	V	N	L	L	M	R	L	T	D	D	Y	L	L	I	T	T	Q	E	N	N	A	V	L	795		
	Sp_Tip1p	757	F	L	N	L	S	L	R	G	F	E	K	H	N	F	S	T	S	L	E	K	T	V	I	N	E	E	N	S	N	G	-	-	-	-	786	
	Sc_Est2p	685	I	K	K	L	A	M	G	G	F	Q	K	Y	N	A	K	A	N	R	D	K	I	L	A	V	S	S	Q	S	D	-	-	-	-	-	713	
	Ea_p123	796	F	I	E	K	L	I	N	V	S	R	E	N	G	F	K	F	N	M	K	K	L	Q	T	S	E	P	L	S	P	S	K	F	A	828		
	Sp_Tip1p	787	-	-	-	I	I	N	N	T	F	F	N	E	S	K	K	R	M	P	F	F	G	F	S	V	N	M	R	S	L	D	T	L	L	816		
	Sc_Est2p	714	-	-	-	D	D	T	V	I	Q	E	C	A	-	-	M	H	I	F	V	K	E	L	E	V	W	K	H	S	S	T	M	739				
	Ea_p123	829	K	Y	G	M	D	S	V	E	E	Q	N	I	V	Q	D	Y	C	D	W	I	G	I	S	I	D	M	K	T	L	A	L	M	P	861		
	Sp_Tip1p	817	A	C	P	K	I	D	E	A	L	F	N	S	T	S	V	E	L	T	K	H	M	G	K	S	F	F	Y	K	I	L	R	S	S	849		
	Sc_Est2p	740	N	N	F	H	I	R	S	K	S	S	K	G	I	F	R	S	L	I	A	L	F	N	T	R	I	S	Y	K	T	I	D	T	N	772		
	Ea_p123	862	N	I	N	L	R	I	E	G	I	L	C	T	L	N	L	N	M	Q	T	K	K	A	S	M	W	L	K	K	L	K	S	F	894			

FIG. 64
(CONTINUED)

A.	Sp_Tlp1p	850	L	A	S	F	A	Q	V	F	I	D	I	T	H	N	S	K	F	N	S	C	C	N	I	Y	R	L	G	Y	S	M	C	M	R	882				
	Sc_Est2p	773	L	N	S	T	N	T	V	L	M	Q	I	D	H	V	V	K	N	I	S	E	C	-	-	-	-	-	-	-	-	-	-	-	-	-	793			
	Ea_p123	895	L	M	N	N	I	T	H	Y	F	R	K	T	I	T	T	E	D	F	A	N	K	T	L	N	K	L	F	I	S	G	G	Y	K	927				
	Sp_Tlp1p	883	A	Q	A	Y	L	K	R	M	K	D	I	F	I	P	Q	R	M	F	I	T	D	L	L	N	V	I	G	R	K	I	W	K	K	915				
	Sc_Est2p	794	-	-	-	Y	K	S	A	F	K	D	L	S	I	N	-	-	V	T	Q	N	M	Q	F	H	S	F	L	Q	R	I	I	E	M	821				
	Ea_p123	928	Y	M	Q	C	A	K	E	Y	K	D	H	E	K	K	N	L	A	M	S	S	M	I	D	L	E	V	S	K	I	I	Y	S	V	960				
	Sp_Tlp1p	916	L	A	E	I	L	G	Y	T	S	R	R	F	L	S	S	A	E	V	K	W	L	F	C	L	G	M	R	D	G	L	K	P	S	948				
	Sc_Est2p	822	T	V	S	G	C	P	I	T	K	C	D	P	L	I	E	Y	E	V	R	F	T	I	L	N	G	F	L	E	S	L	S	S	N	854				
	Ea_p123	961	T	R	A	F	F	K	Y	L	V	C	N	I	K	D	T	I	F	G	E	E	H	Y	P	D	F	F	L	S	T	L	K	H	F	993				
	Sp_Tlp1p	949	F	K	Y	H	P	C	F	E	Q	L	I	Y	Q	F	Q	S	L	T	D	L	I	K	P	L	R	P	V	L	R	Q	V	L	F	981				
	Sc_Est2p	855	T	S	-	-	-	-	-	-	-	-	-	-	-	K	F	K	D	N	I	L	L	R	K	E	I	Q	H	L	Q	A	Y	I	Y	877				
	Ea_p123	994	I	E	I	F	S	-	-	-	-	-	-	-	-	T	K	K	Y	I	F	N	R	V	C	M	I	L	K	A	K	E	A	K	L	K	S	D	Q	C
	Sp_Tlp1p	982	L	H	R	R	I	A	D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	988			
	Sc_Est2p	878	I	Y	I	H	I	V	N	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	884			
	Ea_p123	1024	Q	S	L	I	Q	Y	D	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1031		

FIG. 64
(CONTINUED)

1994

FIG. 64
(CONTINUED)

FIG. 64
(CONTINUED)

B.	Sp_Tip1p	219	WNSISISRFSIF	YR	SS	YK	FK	QD	LY	FN	LH	SI	CD	251
	Sc_Est2p	184	N	-	-	-	-	-	-	-	-	-	-	200
	Ea_p123	218	NEK	-	-	DH	FL	NN	IN	VP	NN	MM	KS	248
	Sp_Tip1p	252	RNTVH	MW	LQ	WIF	PR	QF	GL	IN	AF	VK	QL	284
	Sc_Est2p	201	-	-	-	-	-	-	-	-	-	-	-	223
	Ea_p123	249	R	-	-	-	-	-	-	-	-	-	-	275
	Sp_Tip1p	285	VS	-	-	-	-	-	-	-	-	-	-	313
	Sc_Est2p	224	TN	-	-	-	-	-	-	-	-	-	-	252
	Ea_p123	276	FTN	I	FR	NR	IR	KK	LK	DK	VI	EK	I	308
	Sp_Tip1p	314	LSK	V	YN	HY	CP	Y	ID	-	TH	DD	EK	342
	Sc_Est2p	253	YVS	I	LN	SI	CP	PL	EG	TV	LD	LS	HL	282
	Ea_p123	309	FN	Y	Y	LT	K	SC	PL	PEN	WR	ER	KQ	341
	Sp_Tip1p	343	-	-	-	-	-	-	-	-	-	-	-	359
	Sc_Est2p	283	-	-	-	-	-	-	-	-	-	-	-	299
	Ea_p123	342	SK	Y	Y	E	E	L	F	S	Y	T	T	374
	Sp_Tip1p	360	WGN	QR	I	FE	I	I	L	K	D	LE	T	392
	Sc_Est2p	300	FG	SK	KN	K	G	K	I	I	KN	LN	LL	332
	Ea_p123	375	LT	G	-	R	N	R	K	N	F	Q	K	406
	Sp_Tip1p	393	N	I	K	I	S	E	I	E	W	L	V	425
	Sc_Est2p	333	K	L	R	L	K	D	F	R	W	L	F	362
	Ea_p123	407	K	I	N	T	R	E	I	S	W	M	Q	437

FIG. 64
(CONTINUED)

B.	Sp_Tip1p	426	E	F	I	Y	W	L	Y	N	S	F	I	I	P	I	L	Q	S	FF	Y	I	T	E	S	S	D	L	R	N	R	T	V	Y	458	
	Sc_Est2p	363	C	F	I	S	W	L	F	R	Q	L	I	P	K	I	I	Q	T	FF	Y	C	T	E	I	S	S	T	V	T	-	I	V	Y	394	
	Ea_p123	438	K	L	L	R	W	I	F	E	D	L	V	V	S	L	I	R	C	FF	Y	V	T	E	Q	Q	K	S	Y	S	K	T	Y	Y	470	
	Sp_Tip1p	459	F	R	K	D	I	W	K	L	L	C	R	P	F	I	T	S	M	K	M	E	A	F	E	K	I	N	E	N	N	V	R	M	D	491
	Sc_Est2p	395	F	R	H	D	T	W	N	K	L	I	T	P	F	I	V	E	Y	F	K	T	Y	L	V	E	N	N	V	C	R	N	H	N	S	427
	Ea_p123	471	Y	R	K	N	I	W	D	V	I	M	K	M	S	I	A	D	L	K	K	E	T	L	A	E	V	Q	E	K	E	V	E	E	W	503
	Sp_Tip1p	492	T	Q	K	T	T	L	P	P	A	V	I	R	L	L	P	K	K	-	-	N	T	F	R	L	I	T	N	L	R	K	R	F	L	522
	Sc_Est2p	428	Y	T	L	S	N	F	N	H	S	K	M	R	I	I	P	K	K	S	N	N	E	F	R	I	I	A	I	P	C	R	G	A	D	460
	Ea_p123	504	K	K	S	L	G	F	A	P	G	K	L	R	L	I	P	K	K	-	-	T	T	F	R	P	I	M	T	F	N	K	K	I	V	534
	Sp_Tip1p	523	I	K	M	G	S	N	K	K	M	L	V	S	T	N	Q	T	L	R	P	V	A	S	I	L	K	H	L	I	N	E	-	-	-	552
	Sc_Est2p	461	E	E	E	-	-	F	T	I	Y	K	E	N	H	K	N	A	I	Q	P	T	Q	K	I	L	E	Y	L	R	N	K	R	P	T	491
	Ea_p123	535	N	S	D	-	-	R	K	T	T	K	L	T	T	N	T	K	L	L	N	S	H	L	M	L	K	T	L	K	N	R	-	M	F	564
	Sp_Tip1p	553	E	S	S	G	I	P	F	N	L	E	V	Y	M	K	L	L	T	F	K	K	D	L	L	K	H	R	M	F	G	R	-	K	K	584
	Sc_Est2p	492	S	F	T	K	I	Y	S	P	T	Q	I	A	D	R	I	K	E	F	K	Q	R	L	L	K	K	F	N	N	V	L	P	E	L	524
	Ea_p123	565	K	D	P	F	G	F	A	V	F	N	Y	D	D	V	M	K	K	Y	E	E	F	V	C	K	W	K	Q	V	G	Q	P	K	L	597
	Sp_Tip1p	585	Y	F	V	R	I	D	I	K	S	C	Y	D	R	I	K	Q	D	L	M	F	R	I	V	K	K	K	L	K	D	P	E	-	F	616
	Sc_Est2p	525	Y	F	M	K	F	D	V	K	S	C	Y	D	S	I	P	R	M	E	C	M	R	I	L	K	D	A	L	K	N	E	N	G	F	557
	Ea_p123	598	F	F	A	T	M	D	I	E	K	C	Y	D	S	V	N	R	E	K	L	S	T	F	L	K	T	T	K	L	L	S	S	D	F	630
	Sp_Tip1p	617	V	I	R	K	Y	A	T	I	H	A	T	S	D	R	A	T	K	N	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	634
	Sc_Est2p	558	F	V	R	S	Q	Y	F	F	N	T	N	T	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	570
	Ea_p123	631	W	I	M	T	A	Q	I	L	K	R	K	N	N	I	V	I	D	S	K	N	F	R	K	K	E	M	K	D	Y	F	R	Q	K	663

FIG. 64
(CONTINUED)

B.	Sp_Tip1p	635	FVSEAFSYFDMVPFEK	V	VQLLS	-	-	MKTS	DT	L	FV	665	
	Sc_Est2p	571	-	-	-	-	-	VLKLFN	V	VNASR	-	591	
	Ea_p123	664	FQKIAL	EGGQYPTLFS	V	LENEQNDL	NAKKT	L	I	V	696		
	Sp_Tip1p	666	DFVDYWT	KSSSEIFKML	KEHLS	GHI	VK	IGN	SQ	Y	698		
	Sc_Est2p	592	DNVRTV	HL	SNQDV	IN	VVEME	I	FKT	ALW	VEDKCY	624	
	Ea_p123	697	EAKQRN	YFKKDN	LLQPV	IN	ICQYN	Y	IN	FNGK	FY	729	
	Sp_Tip1p	699	LQKV	G	I	P	Q	G	S	I	L	S	731
	Sc_Est2p	625	I	RED	G	L	F	Q	G	S	S	L	657
	Ea_p123	730	KQTK	G	I	P	Q	G	L	C	V	S	762
	Sp_Tip1p	732	KKG	-	-	-	-	SVLL	RV	V	DD	FL	756
	Sc_Est2p	658	SPSQD	-	-	-	-	TL	IL	KL	AD	FL	684
	Ea_p123	763	DESMN	PEN	P	N	V	N	L	L	M	R	795
	Sp_Tip1p	757	FLNLS	LR	G	F	E	K	H	N	F	S	786
	Sc_Est2p	685	I	KK	L	A	M	G	G	F	Q	K	713
	Ea_p123	796	F	I	E	K	L	I	N	V	S	R	828
	Sp_Tip1p	787	-	-	-	I	I	N	N	T	F	F	816
	Sc_Est2p	714	-	-	-	DD	T	V	I	Q	F	C	739
	Ea_p123	829	K	Y	G	M	D	S	V	E	E	Q	861
	Sp_Tip1p	817	ACP	K	I	D	E	A	L	F	N	S	849
	Sc_Est2p	740	N	N	F	H	I	R	S	K	S	S	772
	Ea_p123	862	N	I	N	L	R	I	E	G	I	L	894

FIG. 64
(CONTINUED)

B.	Sp_Tip1p	850	L	A	S	F	A	Q	V	F	I	D	I	T	H	N	S	K	F	N	S	C	C	N	I	Y	R	L	G	Y	S	M	C	M	R	882		
	Sc_Est2p	773	L	N	S	T	N	T	V	L	M	Q	I	D	H	V	V	K	N	I	S	E	C	-	-	-	-	-	-	-	-	-	-	-	-	-	793	
	Ea_p123	895	L	M	N	N	I	T	H	Y	F	R	K	T	I	T	T	E	D	F	A	N	K	T	L	N	K	L	F	I	S	G	G	Y	K	927		
	Sp_Tip1p	883	A	Q	A	Y	L	K	R	M	K	D	I	F	I	P	Q	R	M	F	I	T	D	L	L	N	V	I	G	R	K	I	W	K	K	915		
	Sc_Est2p	794	-	-	-	Y	K	S	A	F	K	D	L	S	I	N	-	-	V	T	Q	N	M	Q	F	H	S	F	L	Q	R	I	I	E	M	821		
	Ea_p123	928	Y	M	Q	C	A	K	E	Y	K	D	H	F	K	K	N	L	A	M	S	S	M	I	D	L	E	V	S	K	I	I	Y	S	V	960		
	Sp_Tip1p	916	L	A	E	I	L	G	Y	T	S	R	R	F	L	S	S	A	E	V	K	W	L	F	C	L	G	M	R	D	G	L	K	P	S	948		
	Sc_Est2p	822	T	V	S	G	C	P	I	T	K	C	D	P	L	I	E	Y	E	V	R	F	T	I	L	N	G	F	L	E	S	L	S	S	N	854		
	Ea_p123	961	T	R	A	F	F	K	Y	L	V	C	N	I	K	D	T	I	F	G	E	E	H	Y	P	D	F	F	L	S	T	L	K	H	F	993		
	Sp_Tip1p	949	F	K	Y	H	P	C	F	F	E	Q	L	I	Y	Q	F	Q	S	L	T	D	L	I	K	P	L	R	P	V	L	R	Q	V	L	F	981	
	Sc_Est2p	855	T	S	-	-	-	-	-	-	-	-	-	-	-	-	K	F	K	D	N	I	I	L	L	R	K	E	I	Q	H	L	Q	A	Y	I	Y	877
	Ea_p123	994	I	E	I	F	S	-	-	-	-	-	-	-	-	-	T	K	K	Y	I	E	N	R	V	C	M	I	L	K	A	K	E	A	K	L	1023	
	Sp_Tip1p	982	L	H	R	R	I	A	D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	988		
	Sc_Est2p	878	I	Y	I	H	I	V	N	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	884		
	Ea_p123	1024	Q	S	L	I	Q	Y	D	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1031		

FIG. 64
(CONTINUED)

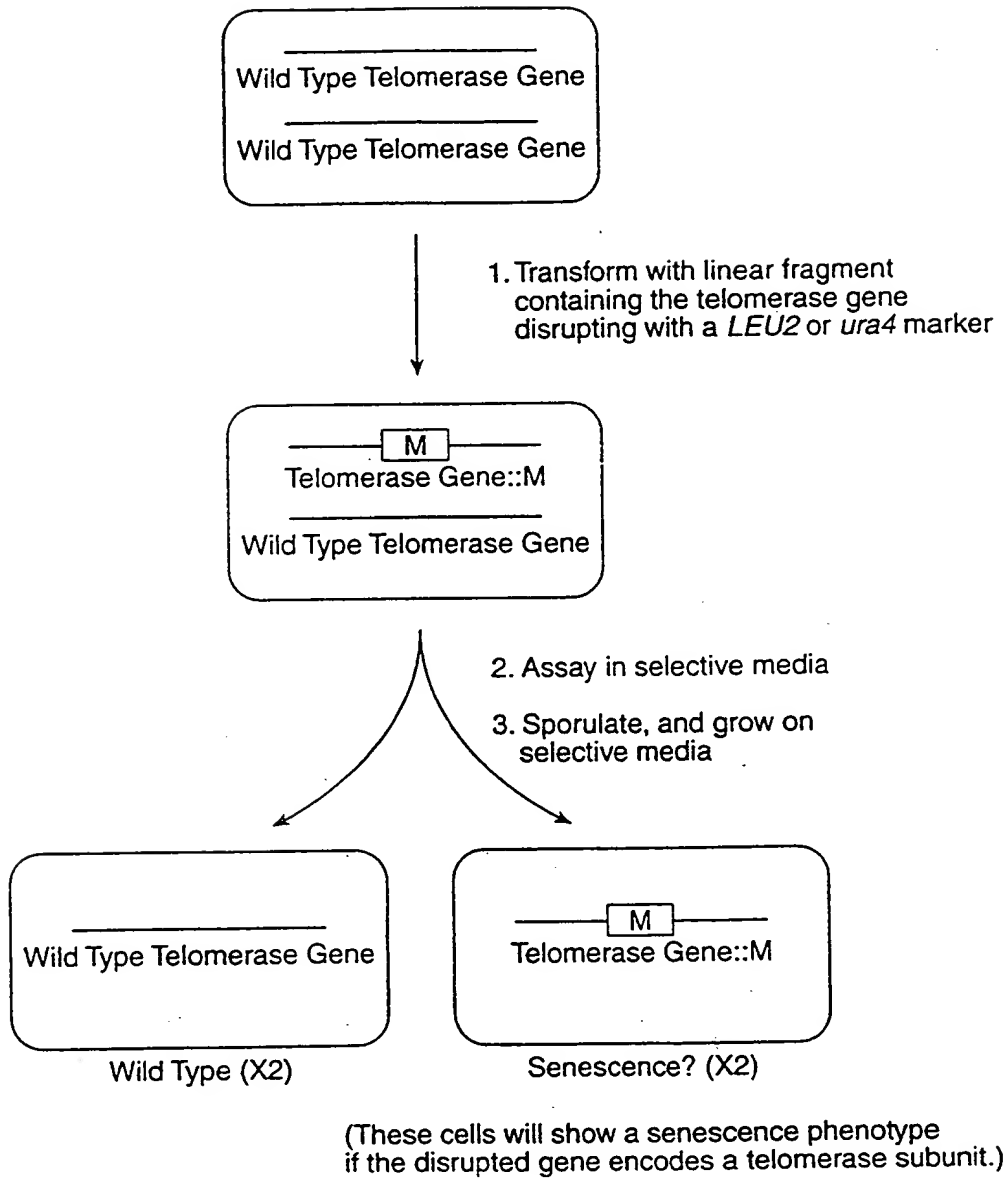
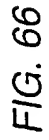


FIG. 65



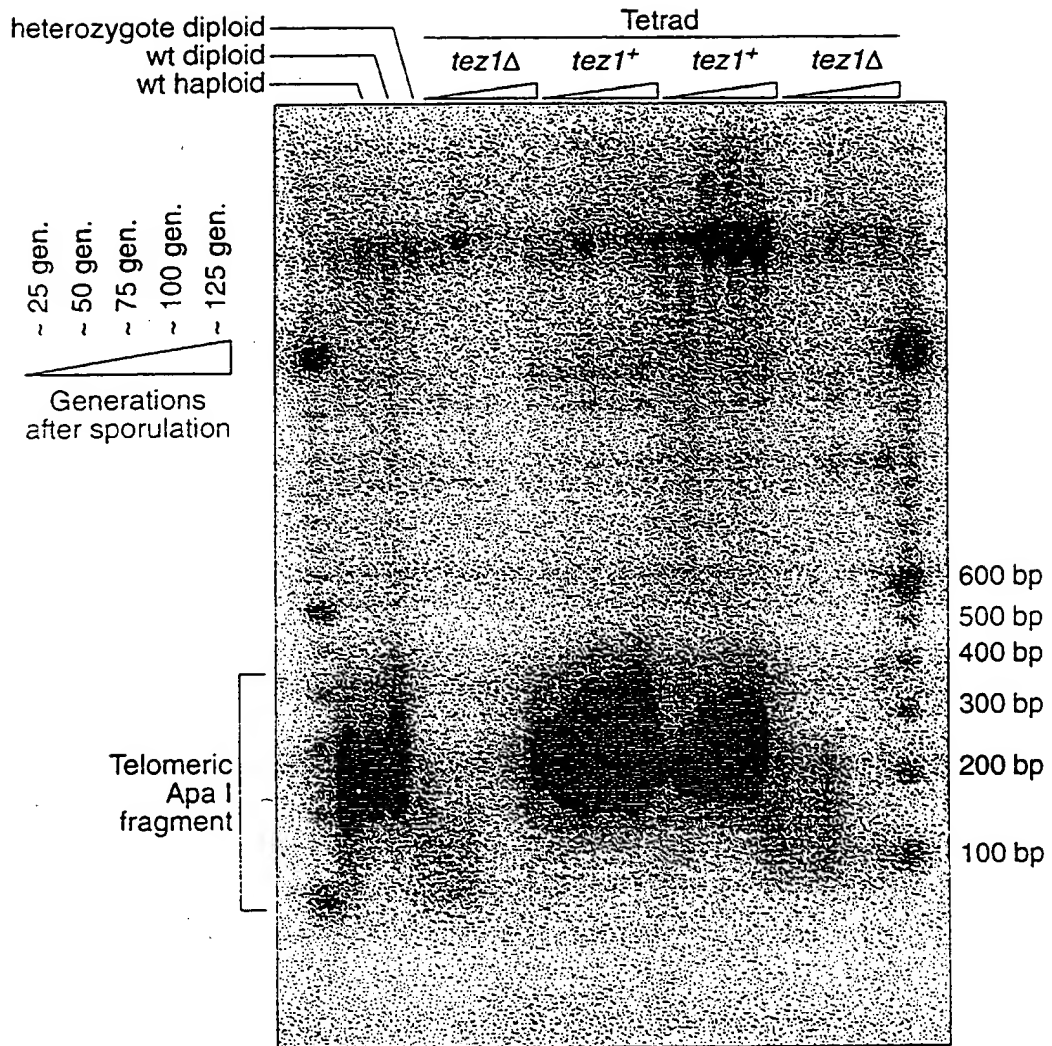


FIG. 67

```

      1
GCCAAGTTCCTGCACTGGCTG  met ser val tyr val val glu leu leu
      ATG AGT GTG TAC GTC GTC GAG CTG CTC

10      20
arg ser phe phe tyr val thr glu thr thr phe gln lys asn arg
AGG TCT TTC TTT TAT GTC ACG GAG ACC ACG TTT CAA AAG AAC AGG

      30
leu phe phe tyr arg lys ser val trp ser lys leu gln ser ile
CTC TTT TTC TAC CGG AAG AGT GTC TGG AGC AAG TTG CAA AGC ATT

40      50
gly ile arg gln his leu lys arg val gln leu arg glu leu ser
GGA ATC AGA CAG CAC TTG AAG ACG GTG CAG CTG CGG GAG CTG TCG

      60
glu ala glu val arg gln his arg glu ala arg pro ala leu leu
GAA GCA GAG GTC AGG CAG CAT CGG GAA GCC AGG CCC GCC CTG CTG

70      80
thr ser arg leu arg phe ile pro lys pro asp gly leu arg pro
ACG TCC AGA CTC CGC TTC ATC CCC AAG CCT GAC GGG CTG CGG CCG

      90
ile val asn met asp tyr val val gly ala arg thr phe arg arg
ATT GTG AAC ATG GAC TAC GTC GTG GGA GCC AGA ACG TTC CGC AGA

100      110
glu lys ala glu arg leu thr ser arg val lys ala leu phe
GAA AAG ARG GCC GAG CGT CTC ACC TCG AGG GTG AAG GCA CTG TTC

      120
ser val leu asn tyr glu arg ala arg arg pro gly leu leu gly
AGC GTG CTC AAC TAC GAG CGG GCG CGG CGC CCC GGC CTC CTG GGC

130      140
ala ser val leu gly leu asp asp ile his arg ala trp arg thr
GCC TCT GTG CTG GGC CTG GAC GAT ATC CAC AGG GCC TGG CGC ACC

      150
phe val leu arg val arg ala gln asp pro pro pro glu leu tyr
TTC GTG CTG CGT GTG CGG GCC CAG GAC CCG CCG CCT GAG CTG TAC

160      170
phe val lys val asp val thr gly ala tyr asp thr ile pro gln
TTT GTC AAG GTG GAT GTG ACG GGC GCG TAC GAC ACC ATC CCC CAG

      180
asp arg leu thr glu val ile ala ser ile ile lys pro gln asn
GAC AGG CTC ACG GAG GTC ATC GCC AGC ATC ATC AAA CCC CAG AAC

190      200
thr tyr cys val arg arg tyr ala val val gln lys ala ala met
ACG TAC TGC GTG CGT CGG TAT GCC GTG GTC CAG AAG GCC GCC ATG

```

FIG. 68

210
 gly thr ser ala arg pro ser arg ala thr ser tyr val gln cys
 GGC ACG TCC GCA AGG CCT TCA AGA GCC ACG TCC TAC GTC CAG TGC

220 230
 gln gly ile pro gln gly ser ile leu ser thr leu leu cys ser
 CAG GGG ATC CCG CAG GGC TCC ATC CTC TCC ACG CTG CTC TGC AGC

240
 leu cys tyr gly asp met glu asn lys leu phe ala gly ile arg
 CTG TGC TAC GGC GAC ATG GAG AAC AAG CTG TTT GCG GGG ATT CGG

250 260
 arg asp gly leu leu leu arg leu val asp asp phe leu leu val
 CGG GAC GGG CTG CTC CTG CGT TTG GTG GAT GAT TTC TTG TTG GTG

270
 thr pro his leu thr his ala lys thr phe leu arg thr leu val
 ACA CCT CAC CTC ACC CAC GCG AAA ACC TTC CTC AGG ACC CTG GTC

280 290
 arg gly val pro glu tyr gly cys val val asn leu arg lys thr
 CGA GGT GTC CCT GAG TAT GGC TGC GTG GTG AAC TTG CGG AAG ACA

300
 val val asn phe pro val glu asp glu ala leu gly gly thr ala
 GTG GTG AAC TTC CCT GTA GAA GAC GAG GCC CTG GGT GGC ACG GCT

310 320
 phe val gln met pro ala his gly leu phe pro trp cys gly leu
 TTT GTT CAG ATG CCG GCC CAC GGC CTA TTC CCC TGG TGC GGC CTG

330
 leu leu asp thr arg thr leu glu val gln ser asp tyr ser ser
 CTG CTG GAT ACC CGG ACC CTG GAG GTG CAG AGC GAC TAC TCC AGC

340 350
 tyr ala arg thr ser ile arg ala ser leu thr phe asn arg gly
 TAT GCC CGG ACC TCC ATC AGA GCC AGT CTC ACC TTC AAC CGC GGC

360
 phe lys ala gly arg asn met arg arg lys leu phe gly val leu
 TTC AAG GCT GGG AGG AAC ATG CGT CGC AAA CTC TTT GGG GTC TTG

370 380
 arg leu lys cys his ser leu phe leu asp leu gln val asn ser
 CGG CTG AAG TGT CAC AGC CTG TTT CTG GAT TTG CAG GTG AAC AGC

390
 leu gln thr val cys thr asn ile tyr lys ile leu leu leu gln
 CTC CAG ACG GTG TGC ACC AAC ATC TAC AAG ATC CTC CTG CTG CAG

400 410
 ala tyr arg phe his ala cys val leu gln leu pro phe his gln
 GCG TAC AGG TTT CAC GCA TGT GTG CTG CAG CTC CCA TTT CAT CAG

FIG. 68
(CONTINUED)

420
 gln val trp lys asn pro his phe ser cys ala ser ser leu thr
 CAA GTT TGG AAG AAC CCA CAT TTT TCC TGC GCG TCA TCT CTG ACA
 430 440
 arg leu pro leu leu leu his pro glu ser gln glu arg arg asp
 CGG CTC CCT CTG CTA CTC CAT CCT GAA AGC CAA GAA CGC AGG GAT
 450
 val ala gly gly gln gly arg arg arg pro ser ala leu arg gly
 GTC GCT GGG GGC CAA GGG CGC CGC CGG CCC TCT GCC CTC CGA GGC
 460 470
 arg ala val ala val pro pro ser ile pro ala gln ala asp ser
 CGT GCA GTG GCT GTG CCA CCA AGC ATT CCT GCT CAA GCT GAC TCG
 480
 thr pro cys his leu arg ala thr pro gly val thr gln asp ser
 ACA CCG TGT CAC CTA CGT GCC ACT CCT GGG GTC ACT CAG GAC AGC
 490 500
 pro asp ala ala glu ser glu ala pro gly asp asp ala asp cys
 CCA GAC GCA GCT GAG TCG GAA GCT CCC GGG GAC GAC GCT GAC TGC
 510
 pro gly gly arg ser gln pro gly thr ala leu arg leu gln asp
 CCT GGA GGC CGC AGC CAA CCC GGC ACT GCC CTC AGA CTT CAA GAC
 520 530
 his pro gly leu met ala thr arg pro gln pro gly arg glu gln
 CAT CCT GGA CTG ATG GCC ACC CGC CCA CAG CCA GGC CGA GAG CAG
 540
 thr pro ala ala leu ser arg arg ala tyr thr ser gln gly gly
 ACA CCA GCA GCC CTG TCA CGC CGG GCT TAT ACG TCC CAG GGA GGG
 550 560
 arg gly gly pro his pro gly leu his arg trp glu ser glu ala
 AGG GGC GGC CCA CAC CCA GGC CTG CAC CGC TGG GAG TCT GAG GCC
 564
 OP
 TGA GTGAGTGTGTTGGCCGAGGCCTGCATGTCCGGCTGAAGGCTGAGTGTCCGGCTGAGGC
 CTGAGCGAGTGTCCAGCCAAGGGCTGAGTGTCCAGCACACCTGCGTTTTCACTTCCCCAC
 AGGCTGGCGTTCGGTCCACCCCAGGGCCAGCTTTTCCTCACCAGGAGCCCGGCTTCCACT
 CCCCACATAGGAATAGTCCATCCCCAGATTGCGCCATTGTTACCCCTTCGCCCTGCCTTCC
 TTTGCCTTCCACCCCCACCATTTCAGGTGGAGACCCTGAGAAGGACCCTGGGAGCTTTGGG
 AATTTGGAGTGACCAAAGGTGTGCCCTGTACACAGGCGAGGACCCTGCACCTGGATGGGG
 GTCCCTGTGGGTCAAATTGGGGGGAGGTGCTGTGGGAGTAAAATACTGAATATATGAGTT
 TTTTCAGTTTTGGAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

Motif -1
 Ep p123 ...LVVSLIRCFYVTEQQKSYSKT...
 Sp Tez1 ...FIIPILQSFFYITESSDLRNRT...
 Sc Est2 ...LIPKIIQTFFYCTEISSTVTIV...
 Hs TCP1 ...YVVELLRSFYVTETTFQKNRL...
 consensus FFY TE

Motif 0
 Ep p123 ...KSLGFAPGKRLRLIPKKT--TFRPIMTFNKKIV...
 Sp Tez1 ...QKTTLPPAVIRLLPKKN--TFRLLITNLRKRFL...
 Sc Est2 ...TLSNFNHSMRIIPKKSNNEFRIIAIPCRGAD...
 Hs TCP1 ...ARPALLTSRLRFIPKPD--GLRPVNM DYVVG...
 consensus R PK R I

Motif A
 Ep p123 ...PKLFFATMDIEKCYDSVNREKLSTFLK...
 Sp Tez1 ...RKKYFVRIDIKSCYDRIKQDLMFRIVK...
 Sc Est2 ...PELYFMKFDVKSCYDSIPRMECMRILK...
 Hs TCP1 ...PELYFVKVDVTGAYDTIPQDRLTEVIA...//...
 consensus F D YD

Motif B
 Ep p123 ...NGKFYKQTKGIPQGLCVSSILSSFYYA...
 Sp Tez1 ...GNSQYLQKVGIPQGSILSSFLCHFyme...
 Sc Est2 ...EDKCYIREDGLFQGSLSAPIVDLVYD...
 Hs TCP1 ...RATSYVQCQGIPOGSILSTLLCSLCYG...
 consensus G QG S

Motif C
 Ep p123 ...PNVNLLMRLTDDYLLITTQENN...
 Sp Tez1 ...KKGSVLLRVVDDFLFITVNKKD...
 Sc Est2 ...SQDTLILKLADDFLIISTDQQQ...
 Hs TCP1 ...RRDGLLLRLVDDFLLVTPHLTH...
 consensus DD L

Motif D
 Ep p123 ...NVSRENGFKFNMKKL...
 Sp Tez1 ...LNLSLRGFEKHNFT...
 Sc Est2 ...KKLAMGGFQKYNKA...
 Hs TCP1 ...LRTLVRGVPEYGCVV...
 consensus G

FIG. 69

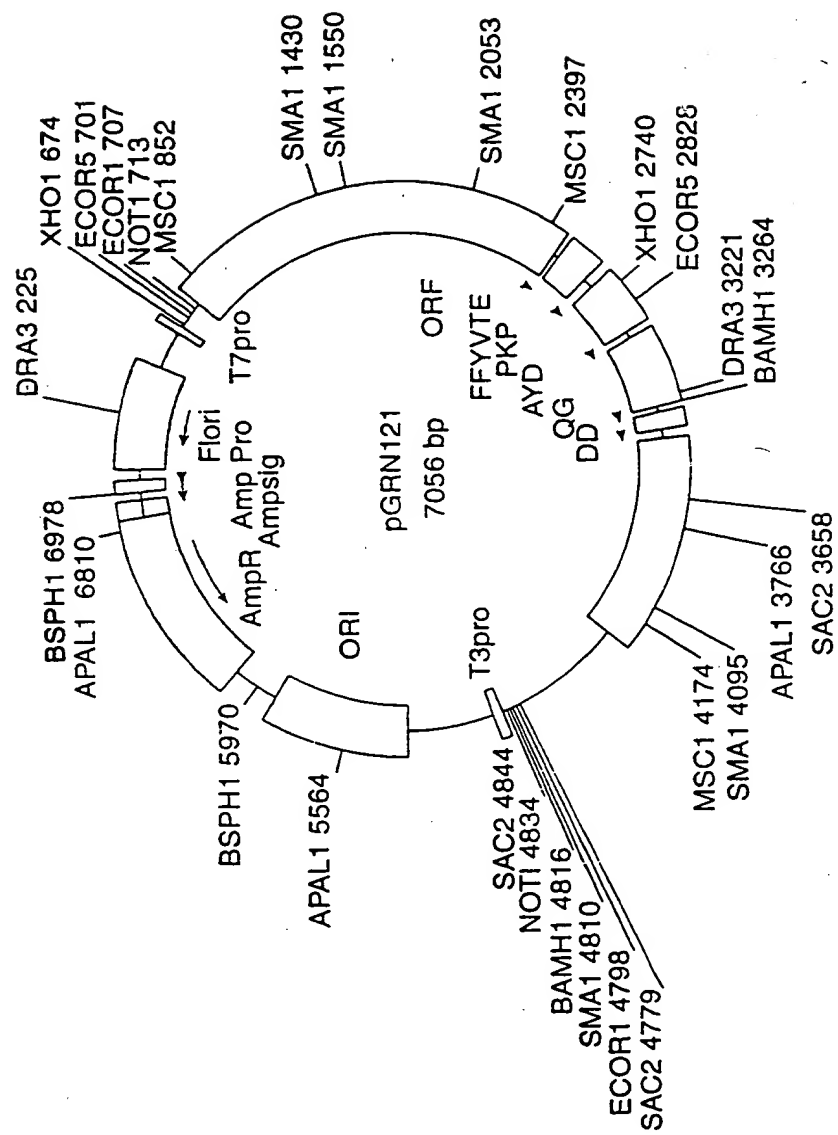


FIG. 70

1	GCAGCGCTGC	GTCCTGCTGC	GCACGTGGGA	AGCCCTGGCC	CCGGCCACCC
51	CCGCGATGCC	GCGCGCTCCC	CGCTGCCGAG	CCGTGCGCTC	CCTGCTGCGC
101	AGCCACTACC	GCGAGGTGCT	GCCGCTGGCC	ACGTTTCGTGC	GGCGCCTGGG
151	GCCCCAGGGC	TGGCGGCTGG	TGCAGCGCGG	GGACCCGGCG	GCTTTCGCG
201	CGNTGGTGCC	CCANTGCNTG	GTGTGCGTGC	CCTGGGANGN	ANGGCNGCCC
251	CCCCCGCCCC	CCTCCTTCCG	CCAGGTGTCC	TGCCTGAANG	ANCTGGTGGC
301	CCGAGTGCTG	CANANGCTGT	GCGANCGCGG	CGCGAANAAC	GTGCTGGCCT
351	TCGGCTTCGC	GCTGCTGGAC	GGGGCCCGCG	GGGGCCCCCC	CGAGGCCTTC
401	ACCACCAGCG	TGCGCAGCTA	CCTGCCCAAC	ACGGTGACCG	ACGCACTGCG
451	GGGGAGCGGG	GCGTGGGGGC	TGCTGCTGCG	CCGCGTGGGC	GACGACGTGC
501	TGGTTCACTT	GCTGGCACGC	TGCGCGNTNT	TTGTGCTGGT	GGNTCCCAGC
551	TGCGCC'ACC	ANGTGTGCGG	GCCGCCGCTG	TACCAGCTCG	GCGCTGCNAC
601	TCAGGCC'CGG	CCCCCGCCAC	ACGCTANTGG	ACCCGAANGC	GTCTGGGATC
651	CAACGGGCCT	GGAACCATAG	CGTCAGGGAG	GCCGGGGTCC	CCCTGGGCTG
701	CCAGCCCCCG	GTGCGAGGAG	GCGCGGGGGC	AGTGCCAGCC	GAAGTCTGCC
751	GTTGCCCAAG	AGGCCCAGGC	GTGGCGCTGC	CCCTGAGCCG	GAGCGGACGC
801	CCGTTGGGCA	GGGGTCTCTG	GCCCCACCCG	GCAGGACGCC	TGGACCGAGT
851	GACCGTGGTT	TCTGTGTGGT	GTCACCTGCC	AGACCCGCCG	AAGAAGCCAC
901	CTCTTTGGAG	GGTGCGCTCT	CTGGCACGCG	CCACTCCCAC	CCATCCGTGG
951	GCCGCCAGCA	CCACGCGGGC	CCCCCATCCA	CATCGCGGCC	ACCACGTCCT
1001	GGGACACGCC	TTGTCCCCCG	GTGTACGCCG	AGACCAAGCA	CTTCCTCTAC
1051	TCCTCAGGCG	ACAAGNACAC	TGCGNCCCTC	CTTCCTACTC	AATATATCTG
1101	AGGCCCAGCC	TGACTGGCGT	TCGGGAGGTT	CGTGAGAGCA	NTCTTTCTGG
1151	TTCCAGGCCT	TGGATGCCAG	GATTCCCCGC	AGGTTGCCCC	GCCTGCCCCA
1201	GCGNTACTGG	CAAATGCGGC	CCCTGTTTCT	GGAGCTGCTT	GGGAACCACG
1251	CGCAGTGCCC	CTACGGGGTG	TTCTCAAGA	CGCACTGCCC	GCTGCGAGCT
1301	GCGGTCACCC	CAGCAGCCGG	TGTCTGTGCC	CGGGAGAAGC	CCCAGGGCTC
1351	TGTGGCGGCC	CCCAGGAGG	AGGAACACAG	ACCCCCGTCG	CCTGGTGCAG
1401	CTGCTCCGCC	AGCACAGCAG	CCCCTGGCAG	GTGTACGGCT	TCGTGCGGGC
1451	CTGCCTGCGC	CGGCTGGTGC	CCCCAGGCCT	CTGGGGCTCC	AGGCACAACG
1501	AACGCCGCTT	CCTCAGGAAC	ACCAAGAAGT	TCATCTCCCT	GGGGAAGCAT
1551	GCCAAGCTCT	CGCTGCAGGA	GCTGACGTGG	AAGATGAGCG	TGCGGGACTG
1601	CGCTTGCGTG	CGCAGGAGCC	CAGGGGTTGG	CTGTGTTCCG	GCCGCAGAGC
1651	ACCGTCTGCG	TGAGGAGATC	CTGGCCAAGT	TCCTGCACTG	GCTGATGAGT
1701	GTGTACGTCG	TCGAGCTGCT	CAGGTCTTTC	TTTTATGTCA	CGGAGACCAC
1751	GTTTCAAAAG	AACAGGCTCT	TTTTCTACCG	GAAGAGTGTC	TGGAGCAAGT
1801	TGCAAAGCAT	TGGAATCAGA	CAGCACTTGA	AGAGGGTGCA	GCTGCGGGAG
1851	CTGTCCGAAG	CAGAGGTCAG	GCAGCATCGG	GAAGCCAGGC	CCGCCCTGCT
1901	GACGTCCAGA	CTCCGCTTCA	TCCCCAAGCC	TGACGGGCTG	CGGCCGATTG
1951	TGAACATGGA	CTACGTCGTG	GGAGCCAGAA	CGTTCCGCAG	AGAAAAGAGG
2001	GCCGAGCGTC	TCACCTCGAG	GGTGAAGGCA	CTGTTTCAGC	TGCTCAACTA
2051	CGAGCGGGCG	CGGCGCCCCG	GCCTCCTGGG	CGCTCTGTG	CTGGGCCTGG
2101	ACGATATCCA	CAGGGCCTGG	CGCACCTTCG	TGCTGCGTGT	GCGGGCCCAG
2151	GACCCGCCGC	CTGAGCTGTA	CTTTGTCAAG	GTGGATGTGA	CGGGCGCGTA
2201	CGACACCATC	CCCCAGGACA	GGCTCACGGA	GGTCATCGCC	AGCATCATCA
2251	AACCCCAGAA	CACGTACTGC	GTGCGTCGGT	ATGCCGTGGT	CCAGAAGGCC
2301	GCCCATGGGC	ACGTCCGCAA	GGCCTTCAAG	AGCCACGTCT	CTACCTTGAC
2351	AGACCTCCAG	CCGTACATGC	GACAGTTCGT	GGCTCACCTG	CAGGANAACA
2401	GCCCGCTGAG	GGATGCCGTC	GTCATCGAGC	AGAGCTCCTC	CCTGAATGAG
2451	GCCAGCAGTG	GCCTCTTCGA	CGTCTTCCTA	CGCTTCATGT	GCCACCACGC

FIG. 71

2501	CGTGCGCATC	AGGGGCAAGT	CCTACGTCCA	GTGCCAGGGG	ATCCCGCAGG
2551	GCTCCATCCT	CTCCACGCTG	CTCTGCAGCC	TGTGCTACGG	CGACATGGAG
2601	AACAAGCTGT	TTGCGGGGAT	TCGGCGGGAC	GGGCTGCTCC	TGCGTTTGGT
2651	GGATGATTTC	TTGTTGGTGA	CACCTCACCT	CACCCACGCG	AAAACCTTCC
2701	TCAGGACCCT	GGTCCGAGGT	GTCCCTGAGT	ATGGCTGCGT	GGTGAACCTG
2751	CGGAAGACAG	TGGTGAACTT	CCCTGTAGAA	GACGAGGCCC	TGGGTGGCAC
2801	GGCTTTTGT	CAGATGCCGG	CCCACGGCCT	ATTCCCCTGG	TGCGGCCTGC
2851	TGCTGGATAC	CCGGACCCTG	GAGGTGCAGA	GCGACTACTC	CAGCTATGCC
2901	CGGACCTCCA	TCAGAGCCAG	TCTCACCTTC	AACCGCGGCT	TCAAGGCTGG
2951	GAGGAACATG	CGTCGCAAAC	TCTTTGGGGT	CTTGCGGCTG	AAGTGTACAC
3001	GCCTGTTTCT	GGATTTGCAG	GTGAACAGCC	TCCAGACGGT	GTGCACCAAC
3051	ATCTACAAGA	TCCTCCTGCT	GCAGGCGTAC	AGGTTTCACG	CATGTGTGCT
3101	GCAGCTCCCA	TTTCATCAGC	AAGTTTGGAA	GAACCCACAC	TTTTTCCTGC
3151	GCGTCATCTC	TGACACGGCC	TCCCTCTGCT	ACTCCATCCT	GAAAGCCAAG
3201	AACGCAGGGA	TGTCGCTGGG	GGCCAAGGGC	GCCGCCGGCC	CTCTGCCCTC
3251	CGAGGCCGTG	CAGTGGCTGT	GCCACCAAGC	ATTCTTGCTC	AAGCTGACTC
3301	GACACCGTGT	CACCTACGTG	CCACTCCTGG	GGTCACTCAG	GACAGCCCAG
3351	ACGCAGCTGA	GTCGGAAGCT	CCCGGGGACG	ACGCTGACTG	CCCTGGAGGC
3401	CGCAGCCAAC	CCGGCACTGC	CCTCAGACTT	CAAGACCATC	CTGGACTGAT
3451	GGCCACCCGC	CCACAGCCAG	GCCGAGAGCA	GACACCAGCA	GCCCTGTCAC
3501	GCCGGGCTCT	ACGTCCCAGG	GAGGGAGGGG	CGGCCCCACAC	CCAGGCCCGC
3551	ACCGCTGGGA	GTCTGAGGCC	TGAGTGAGTG	TTTGGCCGAG	GCCTGCATGT
3601	CCGGCTGAAG	GCTGAGTGTC	CGGCTGAGGC	CTGAGCGAGT	GTCCAGCCAA
3651	GGGCTGAGTG	TCCAGCACAC	CTGCCGTCTT	CACTTCCCCA	CAGGCTGGCG
3701	CTCGGCTCCA	CCCCAGGGCC	AGCTTTTCCT	CACCAGGAGC	CCGGCTTCCA
3751	CTCCCCACAT	AGGAATAGTC	CATCCCCAGA	TTCGCCATTG	TTCACCCCTC
3801	GCCCTGCCCT	CCTTTGCCTT	CCACCCCCAC	CATCCAGGTG	GAGACCCTGA
3851	GAAGGACCCT	GGGAGCTCTG	GGAATTTGGA	GTGACCAAAG	GTGTGCCCTG
3901	TACACAGGCG	AGGACCCTGC	ACCTGGATGG	GGGTCCCTGT	GGGTCAAATT
3951	GGGGGGAGGT	GCTGTGGGAG	TAAAAATACTG	AATATATGAG	TTTTTCAGTT
4001	TTGAAAAAAA	AAAAAAA	AAAAAAA		

FIG. 71
(CONTINUED)

```

GCAGCGCTGCGTCTGCTGCGCACGTGGGAAGCCCTGGCCCCGGCCACCCCGCGATGCC
1 -----+-----+-----+-----+-----+-----+-----+ 60
CGTCGCGACGCGAGGACGACGCGTGACCCCTTCGGGACCGGGCGGTGGGGGCGCTACGG

a   A A L R P A A H V G S P G P G H P R D A -
b   Q R C V L L R T W E A L A P A T P A M P -
c   S A A S C C A R G K P W P R P P P R C R -

GCGCGCTCCCCGCTGCCGAGCCGTGCGCTCCCTGCTGCGCAGCCACTACCGCGAGGTGCT
61 -----+-----+-----+-----+-----+-----+-----+ 120
CGCGCGAGGGGCGACGGCTCGGCACGCGAGGGACGACGCGTCCGGTGATGGCGCTCCACGA

a   A R S P L P S R A L P A A Q P L P R G A -
b   R A P R C R A V R S L L R S H Y R E V L -
c   A L P A A L P C A P C C A A T T A R C C -

GCGGCTGGCCACGTTCG1GCGGCGCCTGGGGCCCCAGGGCTGGCGGCTGGTGACGCGCG
121 -----+-----+-----+-----+-----+-----+-----+ 180
CGGCGACCGGTGCAAGCACGCCGCGGACCCCGGGTCCCGACCGCCGACCACGTTCGCGCC

a   A A G H V R A A P G A P G L A A G A A R -
b   P L A T F V R R L G P Q G W R L V Q R G -
c   R W P R S C G A W G P R A G G W C S A G -

GGACCCGGCGGCTTTCGCGCGNTGGTGCCCCANTGCNTGGTGCGTGCCCTGGGANGN
181 -----+-----+-----+-----+-----+-----+-----+ 240
CCTGGGCGCGCGAAAGGCGCGCNACCACCGGTNACGNACCACACGCACGGGACCCCTNCN

a   G P G G F P R ? G G P ? ? G V R A L G ? -
b   D P A A F R A ? V A ? C ? V C V P W ? ? -
c   T R R L S A R W W P ? A W C A C P G ? ? -

ANGGCNGCCCCCGCGCCCCCTCCTTCCGCCAGGTGTCTGCCTGAANGANCTGGTGCG
241 -----+-----+-----+-----+-----+-----+-----+ 300
TNCCGNCGGGGGCGGCGGGGAGGAAGGCGGTCCACAGGACGGACTTNCNGACCACCG

a   ? A A P R R P L L P P G V L P E ? ? G G -
b   ? ? P P A A P S F R Q V S C L ? ? L V A -
c   G ? P P P P P P S A R C P A * ? ? W W P -

CCGAGTGCTGCANANGCTGTGCGANCGCGGCGGAANAACGTGCTGGCCTTCGGCTTCGC
301 -----+-----+-----+-----+-----+-----+-----+ 360
GGCTCACGACGTNTNCGACACGCTNCGCGCCGCGCTTNTTGCACGACCGGAAGCCGAAGCG

a   P S A A ? A V R ? R R E ? R A G L R L R -
b   R V L ? ? L C ? R G A ? N V L A F G F A -
c   E C C ? ? C A ? A A R ? T C W P S A S R -

GCTGCTGGACGGGGCCCCGCGGGGGCCCCCGAGGCCTTCACCACCAGCGTGCGCAGCTA
361 -----+-----+-----+-----+-----+-----+-----+ 420
CGACGACCTGCCCCGGGCGCCCCCGGGGGGCTCCGGAAGTGGTGGTCCACGCGTCGAT

a   A A G R G P R G P P R G L H H Q R A Q L -
b   L L D G A R G G P P E A F T T S V R S Y -
c   C W T G P A G A P P R P S P P A C A A T -

CCTGCCCAACACGGTGACCGACGCACTGCGGGGAGCGGGGCGTGGGGGCTGCTGCTGCG
421 -----+-----+-----+-----+-----+-----+-----+ 480
GGACGGGTGTGCCACTGGCTGCGTGACGCCCCCTCGCCCCGCACCCCGACGACGACGC

a   P A Q H G D R R T A G E R G V G A A A A -
b   L P N T V T D A L R G S G A W G L L L R -
c   C P T R * P T H C G G A G R G G C C C A -

```

FIG. 72

a b c P R G R R R A G S P A G T L R ? ? C A G -
R V G D D V L V H L L A R C A ? F V L V -
A W A T T C W F T C W H A A R ? L C W W -

541 G G N T C C C A G C T G C G C C T A C C A N G T G T G C G G G C C G C G C T G T A C C A G C T C G G C G C T G C N A C 600
C C N A G G G T C G A C G C G G A T G G T N C A C A C G C C C G G C G G C G A C A T G G T C G A G C C G C G A C G N T G

a b c G S Q L R L P ? V R A A A V P A R R C ? -
? P S C A Y ? V C G P P L Y Q L G A A T -
? P A A P T ? C A G R R C T S S A L ? L -

601 T C A G G C C C G G C C C C C G C C A C A C G C T A N T G G A C C G A A N G C G T C T G G G A T C C A A C G G G C C T 660
A G T C C G G G C C G G G G G C G G T G T G C G A T N A C C T G G G C T T N C C A G A C C C T A G G T T G C C C G G A

a b c S G P A P A T R ? W T R ? R L G S N G P -
Q A R P P P H A ? G P E ? V W D P T G L -
R P G P R H T L ? D P ? A S G I Q R A W -

661 G G A A C C A T A G C G T C A G G G A G G C C G G G T C C C C T G G G C T G C C A G C C C G G G T G C G A G G A G 720
C C T T G G T A T C G A G T C C C T C C G G C C C A G G G G A C C C G A C G G T C G G G G C C C A C G C T C C T C

a b c G T I A S G R P G S P W A A S P G C E E -
E P * R Q G G R G P P G L P A P G A R R -
N H S V R E A G V P L G C Q P R V R G G -

721 G C G C G G G G C A G T G C C A G C C G A A G T C T G C C G T T G C C A A G A G G C C C A G G C G T G G C G C T G C 780
C G C G C C C C G T C A C G G T C G G C T T C A G A C G G C A A C G G T T C T C C G G T C C G A C C G C G A C G

a b c A R G Q C Q P K S A V A Q E A Q A W R C -
R G G S A S R S L P L P K R P R R G A A -
A G A V P A E V C R C P R G P G V A L P -

781 C C C T G A G C C G G A G C G G A C G C C C G T T G G G C A G G G T C C T G G G C C C A C C G G G C A G G A C G C C 840
G G G A C T C G G C C T C G C C T G C G G G C A A C C G T C C C C A G A C C C G G T G G G C C C G T C C T G C G G

a b c P * A G A D A R W A G V L G P P G Q D A -
P E P E R T P V G Q G S W A H P G R T P -
L S R S G R P L G R G P G P T R A G R L -

841 T G G A C C G A G T G A C C G T G G T T T C T G T G T G T C A C C T G C C A G A C C C G C C G A A G A A G C C A C 900
A C C T G G C T C A C T G G C A C C A A G A C A C C A C A G T G G A C G G T C T G G G C G G C T T C T T C G G T G

a b c W T E * P W F L C G V T C Q T R R R S H -
G P S D R G F C V V S P A R P A E E A T -
D R V T V V S V W C H L P D P P K K P P -

901 C T C T T T G G A G G G T C G G C T C T C T G G C A C G C C A C T C C C A C C C A T C C G T G G G C C C C A G C A 960
G A G A A C C T C C C A C G G A G A G A C C G T G C G G G T G A G G T G G G T A G G C A C C C G G C G G T C G T

a b c L F G G C A L W H A P L P P I R G P P A -
S L E G A L S G T R H S H P S V G R Q H -
L W R V R S L A R A T P T H P W A A S T -

961 C C A C G C G G G C C C C C A T C C A C A T C G C G G C C A C C A C G T C C T G G G A C A C G C C T T G T C C C C C G 1020
G G T G C G C C C G G G G G T A G G T G T A G C G C C G T G G T G C A G A C C C T G T G C G G A A C A G G G G C

```

a   P R G P P I H I A A T T S W D T P C P P -
b   H A G P P S T S R P P R P G T R L V P R -
c   T R A P H P H R G H H V L G H A L S P G -

1021 GTGTACGCCGAGACCAAGCACTTCCTCTACTCCTCAGGCGACAAGNACACTGCGNCCCTC
-----+-----+-----+-----+-----+-----+-----+-----+ 1080
CACATGCGGCTCTGGTTTCGTGAAGGAGATGAGGAGTCCGCTGTTTCNTGTGACGCNCGGAG

a   V Y A E T K H F L Y S S G D K ? T A ? L -
b   C T P R P S T S S T P Q A T ? T L R P S -
c   V R R D Q A L P L L L R R Q ? H C ? P P -

1081 CTTCCTACTCAATATATCTGAGGCCAGCCTGACTGGCGTTCCGGAGGTTTCGTGGAGACA
-----+-----+-----+-----+-----+-----+-----+ 1140
GAAGGATGAGTTATATAGACTCCGGGTCCGPACTGACCGCAAGCCCTCCAAGCACCTCTGT

a   L P T Q Y I * G P A * L A F G R F V E T -
b   F L L N I S E A Q P D W R S G G S W R ? -
c   S Y S I Y L R P S L T G V R E V R G D ? -

1141 NTCTTTCTGGTTCCAGGCCTTGGATGCCAGGATTCCCCGCAGGTTGCCCCGCCTGCCCCA
-----+-----+-----+-----+-----+-----+-----+ 1200
NAGAAAGACCAAGGTCCGGAACCTACGGTCTTAAGGGGCGTCCAACGGGGCGGACGGGGT

a   ? F L V P G L G C Q D S P Q V A P P A P -
b   S F W F Q A L D A R I P R R L P R L P Q -
c   L S G S R P W M P G F P A G C P A C P S -

1201 GCGNTACTGGCAAATGCGGCCCTGTTTCTGGAGCTGCTTGGGAACCAACGCGCAGTGCCC
-----+-----+-----+-----+-----+-----+-----+ 1260
CGCNATGACCGTTTACGCCGGGGACAAAGACCTCGACGAACCCCTTGGTGCGCGTCACGGG

a   A ? L A N A A P V S G A A W E P R A V P -
b   R Y W Q M R P L F L E L L G N H A Q C P -
c   ? T G K C G P C F W S C L G T T R S A P -

1261 CTACGGGGTGTTCTCAAGACGCACTGCCCCGCTGCGAGCTGCGGTACCCCCAGCAGCCGG
-----+-----+-----+-----+-----+-----+-----+ 1320
GATGCCCCACAAGGAGTTCTGCGTGACGGGCGACGCTCGACGCCAGTGGGGTTCGTCCGCC

a   L R G V P Q D A L P A A S C G H P S S R -
b   Y G V F L K T H C P L R A A V T P A A G -
c   T G C S S R R T A R C E L R S P Q Q P V -

1321 TGTCTGTGCCCCGGGAGAAGCCCCAGGGCTCTGTGGCGGCCCGAGGAGGAACACAG
-----+-----+-----+-----+-----+-----+-----+ 1380
ACAGACACGGGCCCTCTTCGGGGTCCCGAGACACCGCGGGGGCTCTCTCTCTGTGTGTC

a   C L C P G E A P G L C G G P R G G G T Q -
b   V C A R E K P Q G S V A A P E E E E H R -
c   S V P G R S P R A L W R P P R R R N T D -

1381 ACCCCCGTCGCCTGGTGACGCTGCTCCGCCAGCACAGCAGCCCCCTGGCAGGTGTACGGCT
-----+-----+-----+-----+-----+-----+-----+ 1440
TGGGGGCAGCGGACCAGTCGACGAGGCGGTCTGTCTGTCTGGGGACCGTCCACATGCCGA

a   T P V A W C S C S A S T A A P G R C T A -
b   P P S P G A A A P P A Q Q P L A G V R L -
c   P R R L V Q L L R Q H S S P W Q V Y G F -

1441 TCGTGCGGGCCTGCCTGCGCCGGCTGGTGCCCCCAGGCCTCTGGGGCTCCAGGCACAACG
-----+-----+-----+-----+-----+-----+-----+ 1500
AGCACGCCCCGACGGACGCGGCCGACCACGGGGGTCCGGAACCCGAGGTCCGTGTTGC

```

FIG. 72
(CONTINUED)


```

a   S C G P A C A G W C P Q A S G A P G T T -
b   R A G L P A P A G A P R P L G L Q A Q R -
c   V R A C L R R L V P P G L W G S R H N E -

AACGCCGCTTCCTCAGGAACACCAAGAAGTTTCATCTCCCTGGGGAAGCATGCCAAGCTCT
1501 -----+-----+-----+-----+-----+ 1560
TTGCGGCGAAGGAGTCCTTGTTGGTTCTTCAAGTAGAGGGACCCCTTCGTACGGTTCGAGA

a   N A A S S G T P R S S S P W G S M P S S -
b   T P L P Q E H Q E V H L P G E A C Q A L -
c   R R F L R N T K K F I S L G K H A K L S -

CGCTGCAGGAGCTGACGTGGAAGATGAGCGTGCGGGACTGCGCTTGGCTGCGCAGGAGCC
1561 -----+-----+-----+-----+-----+ 1620
GCGACGTCTCTGACTGCACCTTCTACTCGCACGCCCTGACGCGAACCGACGCGTCTCGG

a   R C R S * R G R * A C G T A L G C A G A -
b   A A G A D V E D E R A G L R L A A Q E P -
c   L Q E L T W K M S V R D C A W L R R S P -

CAGGGGTTGGCTGTGTTCGGGCCGACAGCACCGTCTGCGTGAGGAGATCCTGGCCAAGT
1621 -----+-----+-----+-----+-----+ 1680
GTCCCAACCGACACAAGGCCGGCGTCTCGTGCGACGCACTCCTCTAGGACCGGTTCA

a   Q G L A V F R P Q S T V C V R R S W P S -
b   R G W L C S G R R A P S A * G D P G Q V -
c   G V G C V P A A E H R L R E E I L A K F -

TCCTGCACTGGCTGATGAGTGTGTACGTCTGAGCTGCTCAGGTCTTTCTTTTATGTCA
1681 -----+-----+-----+-----+-----+ 1740
AGGACGTGACCGACTACTCACACATGCAGCAGCTCGACGAGTCCAGAAAGAAAATACAGT

a   S C T G * * V C T S S S C S G L S F M S -
b   P A L A D E C V R R R A A Q V F L L C H -
c   L H W L M S V Y V V E L L R S F F Y V T -

CGGAGACCACGTTTCAAAAGAACAGGCTCTTTTTCTACCGGAAGAGTGTCTGGAGCAAGT
1741 -----+-----+-----+-----+-----+ 1800
GCCTCTGGTGCAAAGTTTCTTGTCGAGAAAAAGATGGCCTTCTCACAGACCTCGTTCA

a   R R P R F K R T G S F S T G R V S G A S -
b   G D H V S K E Q A L F L P E E C L E Q V -
c   E T T F Q K N R L F F Y R K S V W S K L -

TGCAAAGCATTGGAATCAGACAGCACTTGAAGAGGGTGACGTGCGGGAGCTGTCCGAAG
1801 -----+-----+-----+-----+-----+ 1860
ACGTTTCGTAACCTTAGTCTGTCTGAACTTCTCCACGTGACGCCCTCGACAGCCTTC

a   C K A L E S D S T * R G C S C G S C R K -
b   A K H W N Q T A L E E G A A A G A V G S -
c   Q S I G I R Q H L K R V Q L R E L S E A -

CAGAGGTACAGGCAGCATCGGGAAGCCAGGCCCGCCCTGCTGACGTCCAGACTCCGCTTCA
1861 -----+-----+-----+-----+-----+ 1920
GTCTCCAGTCCGTCGTAGCCCTTCGGTCCGGGCGGGACGACTGCAGGTCTGAGGCGAAGT

a   Q R S G S I G K P G P P C * R P D S A S -
b   R G Q A A S G S Q A R P A D V Q T P L H -
c   E V R Q H R E A R P A L L T S R L R F I -

TCCCCAAGCCTGACGGGCTGCGGCCGATTGTGAACATGGACTACGTCTGCGGAGCCAGAA
1921 -----+-----+-----+-----+-----+ 1980
AGGGGTTTCGACTGCCCCGACGCCGGCTAACACTTGTACCTGATGCAGCACCCCTCGGTCTT

```

FIG. 72
(CONTINUED)

[illegible]

FIG. 72
(CONTINUED)

```

a   A R * G M P S S S S R A P P * M R P A V -
b   P A E G C R R H R A E L L P E * G Q Q W -
c   P L R D A V V I E Q S S S L N E A S S G -

2461  GCCTCTTCGACGTCTTCCTACGCTTCATGTGCCACCACGCCGTGCGCATCAGGGGCAAGT
-----+-----+-----+-----+-----+-----+-----+-----+-----+ 2520
      CGGAGAAGCTGCAGAAGGATGCGAAGTACACGGTGGTGCGGCACGCGTAGTCCCCGTTCA

a   A S S T S S Y A S C A T T P C A S G A S -
b   P L R R L P T L H V P P R R A H Q G Q V -
c   L F D V F L R F M C H H A V R I R G K S -

2521  CCTACGTCCAGTGCCAGGGGATCCCGCAGGGCTCCATCCTCTCCACGCTGCTCTGCAGCC
-----+-----+-----+-----+-----+-----+-----+-----+-----+ 2580
      GGATGCAGGTACCGTCCCCCTAGGGCGTCCCGAGGTAGGAGAGGTGCGACGAGACGTCCG

a   P T S S A R G S R R A P S S P R C S A A -
b   L R P V P G D P A G L H P L H A A L Q P -
c   Y V Q C Q G I P Q G S I L S T L L C S L -

2581  TGTGCTACGGCGACATGGAGAACAAGCTGTTTGCGGGGATTGCGCGGGACGGGTGCTCC
-----+-----+-----+-----+-----+-----+-----+-----+-----+ 2640
      ACACGATGCCGCTGTACCTCTTGTTCGACAAACGCCCTAAGCCGCCCTGCCCGACGAGG

a   C A T A T W R T S C L R G F G G T G C S -
b   V L R R H G E Q A V C G D S A G R A A P -
c   C Y G D M E N K L F A G I R R D G L L L -

2641  TCGGTTTGGTGGATGATTTCTTGTGGTGACACCTCACCTCACCCACGCGAAAACCTTCC
-----+-----+-----+-----+-----+-----+-----+-----+-----+ 2700
      ACGCAAACCACCTACTAAAGAACAACCCTGTGGAGTGGAGTGGGTGCGCTTTTGAAGG

a   C V W W M I S C W * H L T S P T R K P S -
b   A F G G * F L V G D T S P H P R E N L P -
c   R L V D D F L L V T P H L T H A K T F L -

2701  TCAGGACCCTGGTCCGAGGTGTCCCTGAGTATGGCTGCGTGGTGAACCTTGCGGAAGACAG
-----+-----+-----+-----+-----+-----+-----+-----+-----+ 2760
      AGTCCITGGGACCAGGCTCCACAGGGACTCATACCGACGCACCACTTGAACGCCTTCTGTC

a   S G P W S E V S L S M A A W * T C G R Q -
b   Q D P G P R C P * V W L R G E L A E D S -
c   R T L V R G V P E Y G C V V N L R K T V -

2761  TGGTGAACCTTCCCTGTAGAAGACGAGGCCCTGGGTGGCACGGCTTTTGTTCAGATGCCGG
-----+-----+-----+-----+-----+-----+-----+-----+-----+ 2820
      ACCACTTGAAGGGACATCTTCTGCTCCGGGACCCACCGTGCCGAAAAACAAGTCTACGGCC

a   W * T S L * K T R P W V A R L L F R C R -
b   G E L P C R R R G P G W H G F C S D A G -
c   V N F P V E D E A L G G T A F V Q M P A -

2821  CCCACGGCCTATTCCCCTGGTGGCGCCTGCTGCTGGATACCCGGACCCTGGAGGTGCAGA
-----+-----+-----+-----+-----+-----+-----+-----+-----+ 2880
      GGGTGCCGGATAAGGGGACCACGCCGACGACCTATGGGCCTGGGACCTCCACGTCT

a   P T A Y S P G A A C C W I P G P W R C R -
b   P R P I P L V R P A A G Y P D P G G A E -
c   H G L F P W C G L L L D T R T L E V Q S -

2881  GCGACTACTCCAGCTATGCCCGGACCTCCATCAGAGCCAGTCTCACCTTCAACCGCGGCT
-----+-----+-----+-----+-----+-----+-----+-----+-----+ 2940
      CGCTGATGAGGTGCATACGGGCTGGAGGTAGTCTCGGTCAGAGTGAAGTTGGCGCCGA

```

FIG. 72
(CONTINUED)

94/103

```

a   A T T P A M P G P P S E P V S P S T A A -
b   R L L Q L C P D L H Q S Q S H L Q P R L -
c   D Y S S Y A R T S I R A S L T F N R G F -

TCAAGGCTGGGAGGAACATGCGTCGCAAACTCTTTGGGGTCTTGGGGCTGAAGTGTACACA
2941 -----+-----+-----+-----+-----+ 3000
AGTTCCGACCCCTCCTTGTACGCAGCGTTTGAGAAACCCCAAGCCGACTTCACAGTGT

a   S R L G G T C V A N S L G S C G * S V T -
b   Q G W E E H A S Q T L W G L A A E V S Q -
c   K A G R N M R R K L F G V L R L K C H S -

GCCTGTTTCTGGATTTCAGGTGAACAGCCCTCCAGACGGTGTGCACCAACATCTACAAGA
3001 -----+-----+-----+-----+-----+ 3060
CGGACAAAGACCTAAACGTCCACTTGTGCGAGGTCTGCCACACGTGGTTGTAGATCTTCT

a   A C F W I C R * T A S R R C A P T S T R -
b   P V S G F A G E Q P P D G V H Q H L Q D -
c   L F L D L Q V N S L Q T V C T N I Y K I -

TCCTCCTGCTGCAGGCGTACAGGTTTCACGCATGTGTGCTGCAGCTCCCATTTTCATCAGC
3061 -----+-----+-----+-----+-----+ 3120
AGGAGGACGACGTCCGCATGTCCAAAGTGGTACACACGACGTGAGGGTAAAGTAGTCG

a   S S C C R R T G F T H V C C S S H F I S -
b   P P A A G V Q V S R M C A A A P I S S A -
c   L L L Q A Y R F H A C V L Q L P F H Q Q -

AAGTTTGGAAGAACCCACATTTTCTGCGCGTCATCTCTGACACGGCCTCCCTCTGCT
3121 -----+-----+-----+-----+-----+ 3180
TTCAAACCTTCTTGGGGTGTAAAAAGGACGCGCAGTAGAGACTGTGCCGGAGGGAGACGA

a   K F G R T P H F S C A S S L T R P P S A -
b   S L E E P H I F P A R H L * H G L P L L -
c   V W K N P T F F L R V I S D T A S L C Y -

ACTCCATCCTGAAAGCCAAGAACGCAGGGATGTGCTGGGGGCCAAGGGCGCCGCCGGCC
3181 -----+-----+-----+-----+-----+ 3240
TGAGGTAGGACTTTTCGGTTCTTTCGCTCCCTACAGCGACCCCCGGTTCCCGCGGCGGCCG

a   T P S * K P R T Q G C R W G P R A P P A -
b   L H P E S Q E R R D V A G G Q G R R R P -
c   S I L K A K N A G M S L G A K G A A G P -

CTCTGCCCTCCGAGGCCGTGCAGTGGCTGTGCCACCAAGCATTCCTGCTCAAGCTGACTC
3241 -----+-----+-----+-----+-----+ 3300
GAGACGGGAGGCTCCGGCACGTACCGACACGGTGGTTGCTAAGGACGAGTTCGACTGAG

a   L C P P R P C S G C A T K H S C S S * L -
b   S A L R G R A V A V P P S I P A Q A D S -
c   L P S E A V Q W L C H Q A F L L K L T R -

GACACCGTGTACCTACGTGCCACTCCTGGGGTCACTCAGGACAGCCCAGACGCAGCTGA
3301 -----+-----+-----+-----+-----+ 3360
CTGTGGCACAGTGGATGCACGGTGAGGACCCAGTGAGTCTGTGCGGTCTGCGTCACT

a   D T V S P T C H S W G H S G Q P R R S * -
b   T P C H L R A T P G V T Q D S P D A A E -
c   H R V T Y V P L L G S L R T A Q T Q L S -

GTCCGAAGCTCCCGGGGACGACGCTGACTGCCCTGGAGGCCGAGCCAACCCGGCACTGC
3361 -----+-----+-----+-----+-----+ 3420
CAGCCTTCGAGGGCCCTGCTGCGACTGACGGGACCTCCGGCGTCCGTTGGGCGGTGACG

```

FIG. 72
(CONTINUED)

```

a   V G S S R G R R * L P W R P Q P T R H C -
b   S E A P G D D A D C P G G R S Q P G T A -
c   R K L P G T T L T A L E A A A N P A L P -

CCTCAGACTTCAAGACCATCCTGGACTGATGGCCACCCGCCACAGCCAGGCCGAGAGCA
3421 -----+-----+-----+-----+-----+ 3480
GGAGTCTGAAGTTCTGGTAGGACCTGACTACCGGTGGGCGGGTGTCTGGTCCGGCTCTCGT

a   P Q T S R P S W T D G H P P T A R P R A -
b   L R L Q D H P G L M A T R P Q P G R E Q -
c   S D F K T I L D * W P P A H S Q A E S R -

GACACCAGCAGCCCTGTCACGCCGGGCTCTACGTCCCAGGGAGGGAGGGGCGGCCACAC
3481 -----+-----+-----+-----+-----+ 3540
CTGTGGTCTGTCGGGACAGTGCAGGCCGAGATGCAGGGTCCCTCCCTCCCGCCGGGTGTG

a   D T S S P V T P G S T S Q G G R G G P H -
b   T P A A L S R R A L R P R E G G A A H T -
c   H Q Q P C H A G L Y V P G R E G R P T P -

CCAGGCCCGCACCGCTGGGAGTCTGAGGCCTGAGTGAGTGTMTGGCCGAGGCCTGCATGT
3541 -----+-----+-----+-----+-----+ 3600
GGTCCGGGCGTGCGGACCCCTCAGACTCCGGACTCACTCACAACCGGCTCCGGACGTACA

a   P G P H R W E S E A * V S V W P R P A C -
b   Q A R T A G S L R P E * V F G R G L H V -
c   R P A P L G V * G L S E C L A E A C M S -

CCGGCTGAAGGCTGAGTGTCCGGCTGAGGCCTGAGCGAGTGTCCAGCCAAGGGCTGAGTG
3601 -----+-----+-----+-----+-----+ 3660
GGCCGACTTCCGACTCACAGGCCGACTCCGGACTCGCTCACAGGTGGTTCCTCCGACTCAC

a   P A E G * V S G * G L S E C P A K G * V -
b   R L K A E C P A E A * A S V Q P R A E C -
c   G * R L S V R L R P E R V S S Q G L S V -

TCCAGCACACCTGCCGTCTTCACTTCCCCACAGGCTGGCGCTCGGCTCCACCCCAGGGCC
3661 -----+-----+-----+-----+-----+ 3720
AGGTCTGTGTGGACGGCAGAAGTGAAGGGGTGTCCGACCGCGAGCCGAGGTGGGGTCCCGG

a   S S T P A V F T S P Q A G A R L H P R A -
b   P A H L P S S L P H R L A L G S T P G P -
c   Q H T C R L H F P T G W R S A P P Q G Q -

AGCTTTTCTCACCAGGAGCCCGGCTTCCACTCCCCACATAGGAATAGTCCATCCCCAGA
3721 -----+-----+-----+-----+-----+ 3780
TCGAAAAGGAGTGGTCTCGGGCCGAAGGTGAGGGGTGTATCCTTATCAGGTAGGGGTCT

a   S F S S P G A R L P L P T * E * S I P R -
b   A F P H Q E P G F H S P H R N S P S P D -
c   L F L T R S P A S T P H I G I V H P Q I -

TTCGCCATTGTTCACCCCTCGCCCTGCCCTCTTTGCCCTCCACCCCCACCATCCAGGTG
3781 -----+-----+-----+-----+-----+ 3840
AAGCGGTAACAAGTGGGGAGCGGGACGGGAGGAAACGGAAGGTGGGGGTGGTAGGTCCAC

a   F A I V H P S P C P P L P S T P T I Q V -
b   S P L F T P R P A L L C L P P P P S R W -
c   R H C S P L A L P S F A F H P H H P G G -

GAGACCTGAGAAGGACCCTGGGAGCTCTGGGAATTTGGAGTGACCAAAGGTGTGCCCTG
3841 -----+-----+-----+-----+-----+ 3900
CTCTGGGACTCTTCTCGGACCCCTCGAGACCCCTTAAACCTCACTGGTTTCCACACGGGAC

```

FIG. 72
(CONTINUED)

```

a   E T L R R T L G A L G I W S D Q R C A L -
b   R P * E G P W E L W E F G V T K G V P C -
c   D P E K D P G S S G N L E * P K V C P V -

TACACAGGCGAGGACCCCTGCACCTGGATGGGGGTCCCTGTGGGTCAAATTGGGGGGAGGT
3901 -----+-----+-----+-----+-----+ 3960
ATGTGTCCGCTCCTGGGACGTGGACCTACCCCCAGGGACACCCAGTTTAACCCCCCTCCA

a   Y T G E D P A P G W G S L W V K L G G G -
b   T Q A R T L H L D G G P C G S N W G E V -
c   H R R G P C T W M G V P V G Q I G G R C -

GCTGTGGGAGTAAAATACTGAATATATGAGTTTTTTCAGTTTTTGAAAAAAAAAAAAAAAA
3961 -----+-----+-----+-----+-----+ 4020
CGACACCCCTCATTTTATGACTTATATACTCAAAAAGTCAAAACTTTTTTTTTTTTTTTTTT

a   A V G V K Y * I Y E F F S F E K K K K K -
b   L W E * N T E Y M S F S V L K K K K K K -
c   C G S K I L N I * V F Q F * K K K K K K -

AAAAAAAAA
4021 ----- 4029
TTTTTTTTT

a   K K K -
b   K K -
c   K K -

```

FIG. 72
(CONTINUED)

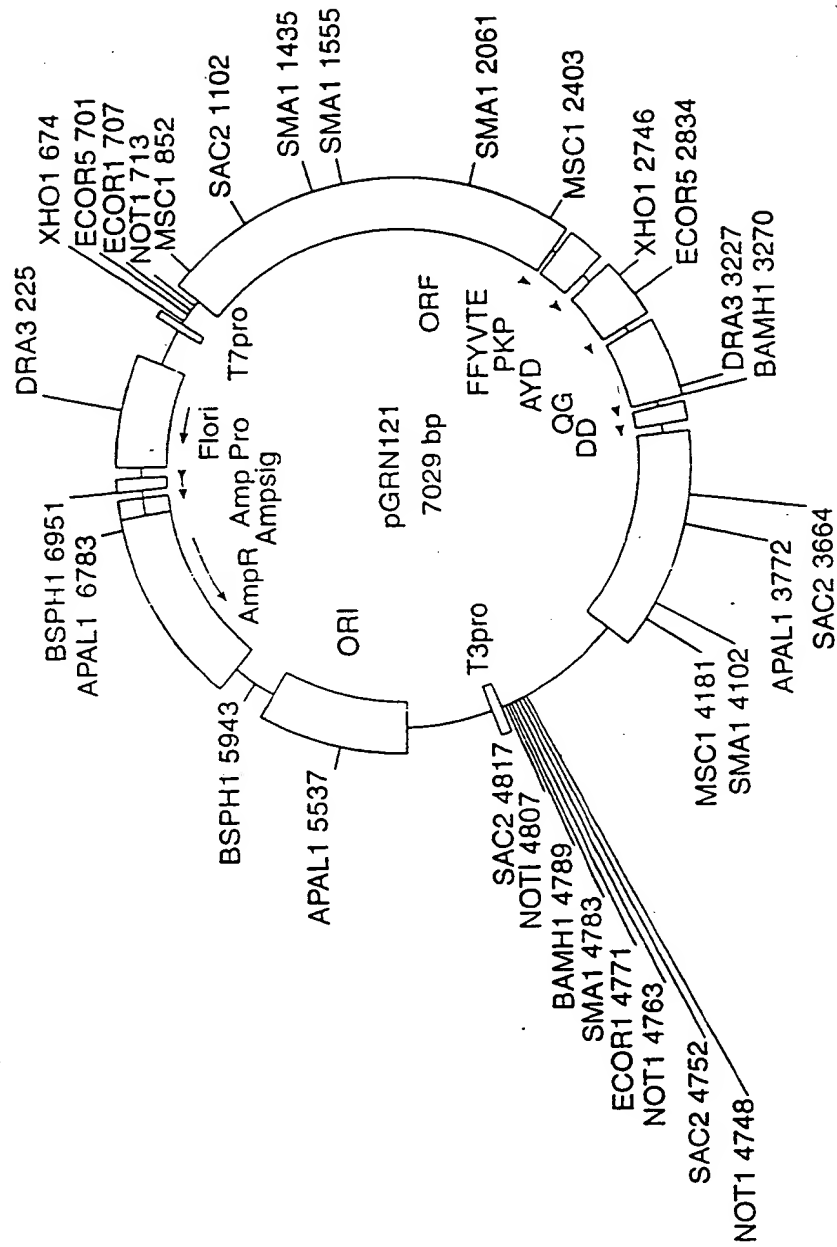


FIG. 73

GCAGCGCTGCGTCCTGCTGCGCACGTGGGAAGCCCTGGCCCCGGCCACCCCCGCG														1 met ATG
														10
pro	arg	ala	pro	arg	cys	arg	ala	val	arg	ser	leu	leu	arg	ser
CCG	CGC	GCT	CCC	CGC	TGC	CGA	GCC	GTG	CGC	TCC	CTG	CTG	CGC	AGC
														20
his	tyr	arg	glu	val	leu	pro	leu	ala	thr	phe	val	arg	arg	leu
CAC	TAC	CGC	GAG	GTG	CTG	CCG	CTG	GCC	ACG	TTC	GTG	CGG	CGC	CTG
														30
gly	pro	gln	gly	trp	arg	leu	val	gln	arg	gly	asp	pro	ala	ala
GGG	CCC	CAG	GGC	TGG	CGG	CTG	GTG	CAG	CGC	GGG	GAC	CCG	GCG	GCT
														40
phe	arg	ala	leu	val	ala	gln	cys	leu	val	cys	val	pro	trp	asp
TTC	CGC	GCG	CTG	GTG	GCC	CAG	TGC	CTG	GTG	TGC	GTG	CCC	TGG	GAC
														50
ala	arg	pro	pro	pro	ala	ala	pro	ser	phe	arg	gln	val	ser	cys
GCA	CGG	CCG	CCC	CCC	GCC	GCC	CCC	TCC	TTC	CGC	CAG	GTG	TCC	TGC
														60
leu	lys	glu	leu	val	ala	arg	val	leu	gln	arg	leu	cys	glu	arg
CTG	AAG	GAG	CTG	GTG	GCC	CGA	GTG	CTG	CAG	AGG	CTG	TGC	GAG	CGC
														70
gly	ala	lys	asn	val	leu	ala	phe	gly	phe	ala	leu	leu	asp	gly
GGC	GCG	AAG	AAC	GTG	CTG	GCC	TTC	GGC	TTC	GCG	CTG	CTG	GAC	GGG
														80
ala	arg	gly	gly	pro	pro	glu	ala	phe	thr	thr	ser	val	arg	ser
GCC	CGC	GGG	GGC	CCC	CCC	GAG	GCC	TTC	ACC	ACC	AGC	GTG	CGC	AGC
														90
tyr	leu	pro	asn	thr	val	thr	asp	ala	leu	arg	gly	ser	gly	ala
TAC	CTG	CCC	AAC	ACG	GTG	ACC	GAC	GCA	CTG	CGG	GGG	AGC	GGG	GCG
														100
trp	gly	leu	leu	leu	arg	arg	val	gly	asp	asp	val	leu	val	his
TGG	GGG	CTG	CTG	CTG	CGC	CGC	GTG	GGC	GAC	GAC	GTG	CTG	GTT	CAC
														110
leu	leu	ala	arg	cys	ala	leu	phe	val	leu	val	ala	pro	ser	cys
CTG	CTG	GCA	CGC	TGC	GCG	CTC	TTT	GTG	CTG	GTG	GCT	CCC	AGC	TGC
														120
ala	tyr	gln	val	cys	gly	pro	pro	leu	tyr	gln	leu	gly	ala	ala
GCC	TAC	CAG	GTG	TGC	GGG	CCG	CCG	CTG	TAC	CAG	CTC	GGC	GCT	GCC
														130
thr	gln	ala	arg	pro	pro	pro	his	ala	ser	gly	pro	arg	arg	arg
ACT	CAG	GCC	CGG	CCC	CCG	CCA	CAC	GCT	AGT	GGA	CCC	CGA	AGG	CGT
														140
														150
														160
														170
														180
														190

FIG. 74

200												210			
leu	gly	cys	glu	arg	ala	trp	asn	his	ser	val	arg	glu	ala	gly	
CTG	GGA	TGC	GAA	CGG	GCC	TGG	AAC	CAT	AGC	GTC	AGG	GAG	GCC	GGG	
220															
val	pro	leu	gly	leu	pro	ala	pro	gly	ala	arg	arg	arg	gly	gly	
GTC	CCC	CTG	GGC	CTG	CCA	GCC	CCG	GGT	GCG	AGG	AGG	CGC	GGG	GGC	
230												240			
ser	ala	ser	arg	ser	leu	pro	leu	pro	lys	arg	pro	arg	arg	gly	
AGT	GCC	AGC	CGA	AGT	CTG	CCG	TTG	CCC	AAG	AGG	CCC	AGG	CGT	GGC	
250															
ala	ala	pro	glu	pro	glu	arg	thr	pro	val	gly	gln	gly	ser	trp	
GCT	GCC	CCT	GAG	CCG	GAG	CGG	ACG	CCC	GTT	GGG	CAG	GGG	TCC	TGG	
260												270			
ala	his	pro	gly	arg	thr	arg	gly	pro	ser	asp	arg	gly	phe	cys	
GCC	CAC	CCG	GGC	AGG	ACG	CGT	GGA	CCG	AGT	GAC	CGT	GGT	TTC	TGT	
280															
val	val	ser	pro	ala	arg	pro	ala	glu	glu	ala	thr	ser	leu	glu	
GTG	GTG	TCA	CCT	GCC	AGA	CCC	GCC	GAA	GAA	GCC	ACC	TCT	TTG	GAG	
290												300			
gly	ala	leu	ser	gly	thr	arg	his	ser	his	pro	ser	val	gly	arg	
GGT	GCG	CTC	TCT	GGC	ACG	CGC	CAC	TCC	CAC	CCA	TCC	GTG	GGC	CGC	
310															
gln	his	his	ala	gly	pro	pro	ser	thr	ser	arg	pro	pro	arg	pro	
CAG	CAC	CAC	GCG	GGC	CCC	CCA	TCC	ACA	TCG	CGG	CCA	CCA	CGT	CCC	
320												330			
trp	asp	thr	pro	cys	pro	pro	val	tyr	ala	glu	thr	lys	his	phe	
TGG	GAC	ACG	CCT	TGT	CCC	CCG	GTG	TAC	GCC	GAG	ACC	AAG	CAC	TTC	
340															
leu	tyr	ser	ser	gly	asp	lys	glu	gln	leu	arg	pro	ser	phe	leu	
CTC	TAC	TCC	TCA	GGC	GAC	AAG	GAG	CAG	CTG	CGG	CCC	TCC	TTC	CTA	
350												360			
leu	ser	ser	leu	arg	pro	ser	leu	thr	gly	ala	arg	arg	leu	val	
CTC	AGC	TCT	CTG	AGG	CCC	AGC	CTG	ACT	GGC	GCT	CGG	AGG	CTC	GTG	
370															
glu	thr	ile	phe	leu	gly	ser	arg	pro	trp	met	pro	gly	thr	pro	
GAG	ACC	ATC	TTT	CTG	GGT	TCC	AGG	CCC	TGG	ATG	CCA	GGG	ACT	CCC	
380												390			
arg	arg	leu	pro	arg	leu	pro	gln	arg	tyr	trp	gln	met	arg	pro	
CGC	AGG	TTG	CCC	CGC	CTG	CCC	CAG	CGC	TAC	TGG	CAA	ATG	CGG	CCC	
400															
leu	phe	leu	glu	leu	leu	gly	asn	his	ala	gln	cys	pro	tyr	gly	
CTG	TTT	CTG	GAG	CTG	CTT	GGG	AAC	CAC	GCG	CAG	TGC	CCC	TAC	GGG	
410												420			
val	leu	leu	lys	thr	his	cys	pro	leu	arg	ala	ala	val	thr	pro	
GTG	CTC	CTC	AAG	ACG	CAC	TGC	CCG	CTG	CGA	GCT	GCG	GTC	ACC	CCA	

FIG. 74
(CONTINUED)

100/103

```

                                430
ala ala gly val cys ala arg glu lys pro gln gly ser val ala
GCA GCC GGT GTC TGT GCC CGG GAG AAG CCC CAG GGC TCT GTG GCG

                                440
ala pro glu glu glu asp thr asp pro arg arg leu val gln leu
GCC CCC GAG GAG GAG GAC ACA GAC CCC CGT CGC CTG GTG CAG CTG

                                450
                                460
leu arg gln his ser ser pro trp gln val tyr gly phe val arg
CTC CGC CAG CAC AGC AGC CCC TGG CAG GTG TAC GGC TTC GTG CGG

                                470
                                480
ala cys leu arg arg leu val pro pro gly leu trp gly ser arg
GCC TGC CTG CGC CGG CTG GTG CCC CCA GGC CTC TGG GGC TCC AGG

                                490
his asn glu arg arg phe leu arg asn thr lys lys phe ile ser
CAC AAC GAA CGC CGC TTC CTC AGG AAC ACC AAG AAG TTC ATC TCC

                                500
                                510
leu gly lys his ala lys leu ser leu gln glu leu thr trp lys
CTG GGG AAG CAT GCC AAG CTC TCG CTG CAG GAG CTG ACG TGG AAG

                                520
met ser val arg asp cys ala trp leu arg arg ser pro gly val
ATG AGC GTG CGG GAC TGC GCT TGG CTG CGC AGG AGC CCA GGG GTT

                                530
                                540
gly cys val pro ala ala glu his arg leu arg glu glu ile leu
GGC TGT GTT CCG GCC GCA GAG CAC CGT CTG CGT GAG GAG ATC CTG

                                550
ala lys phe leu his trp leu met ser val tyr val val glu leu
GCC AAG TTC CTG CAC TGG CTG ATG AGT GTG TAC GTC GTC GAG CTG

                                560
                                570
leu arg ser phe phe tyr val thr glu thr thr phe gln lys asn
CTC AGG TCT TTC TTT TAT GTC ACG GAG ACC ACG TTT CAA AAG AAC

                                580
arg leu phe phe tyr arg pro ser val trp ser lys leu gln ser
AGG CTC TTT TTC TAC CGG CCG AGT GTC TGG AGC AAG TTG CAA AGC

                                590
                                600
ile gly ile arg gln his leu lys arg val gln leu arg glu leu
ATT GGA ATC AGA CAG CAC TTG AAG AGG GTG CAG CTG CGG GAG CTG

                                610
ser glu ala glu val arg gln his arg glu ala arg pro ala leu
TCG GAA GCA GAG GTC AGG CAG CAT CGG GAA GCC AGG CCC GCC CTG

                                620
                                630
leu thr ser arg leu arg phe ile pro lys pro asp gly leu arg
CTG ACG TCC AGA CTC CGC TTC ATC CCC AAG CCT GAC GGG CTG CGG

                                640
pro ile val asn met asp tyr val val gly ala arg thr phe arg
CCG ATT GTG AAC ATG GAC TAC GTC GTG GGA GCC AGA ACG TTC CGC

```

FIG. 74

(CONTINUED)

101/103

650 arg glu lys arg ala glu arg leu thr ser arg val lys ala leu
 AGA GAA AAG AGG GCC GAG CGT CTC ACC TCG AGG GTG AAG GCA CTG

660
 phe ser val leu asn tyr glu arg ala arg arg pro gly leu leu
 TTC AGC GTG CTC AAC TAC GAG CGG GCG CGG CGC CCC GGC CTC CTG

670
 gly ala ser val leu gly leu asp asp ile his arg ala trp arg
 GGC GCC TCT GTG CTG GGC CTG GAC GAT ATC CAC AGG GCC TGG CGC

680
 thr phe val leu arg val arg ala gln asp pro pro pro glu leu
 ACC TTC GTG CTG CGT GTG CGG GCC CAG GAC CCG CCG CCT GAG CTG

690
 tyr phe val lys val asp val thr gly ala tyr asp thr ile pro
 TAC TTT GTC AAG GTG GAT GTG ACG GGC GCG TAC GAC ACC ATC CCC

700
 gln asp arg leu thr glu val ile ala ser ile ile lys pro gln
 CAG GAC AGG CTC ACG GAG GTC ATC GCC AGC ATC ATC AAA CCC CAG

710
 asn thr tyr cys val arg arg tyr ala val val gln lys ala ala
 AAC ACG TAC TGC GTG CGT CGG TAT GCC GTG GTC CAG AAG GCC GCC

720
 his gly his val arg lys ala phe lys ser his val ser thr leu
 CAT GGG CAC GTC CGC AAG GCC TTC AAG AGC CAC GTC TCT ACC TTG

730
 thr asp leu gln pro tyr met arg gln phe val ala his leu gln
 ACA GAC CTC CAG CCG TAC ATG CGA CAG TTC GTG GCT CAC CTG CAG

740
 glu thr ser pro leu arg asp ala val val ile glu gln ser ser
 GAG ACC AGC CCG CTG AGG GAT GCC GTC GTC ATC GAG CAG AGC TCC

750
 ser leu asn glu ala ser ser gly leu phe asp val phe leu arg
 TCC CTG AAT GAG GCC AGC AGT GGC CTC TTC GAC GTC TTC CTA CGC

760
 phe met cys his his ala val arg ile arg gly lys ser tyr val
 TTC ATG TGC CAC CAC GCC GTG CGC ATC AGG GGC AAG TCC TAC GTC

770
 gln cys gln gly ile pro gln gly ser ile leu ser thr leu leu
 CAG TGC CAG GGG ATC CCG CAG GGC TCC ATC CTC TCC ACG CTG CTC

780
 cys ser leu cys tyr gly asp met glu asn lys leu phe ala gly
 TGC AGC CTG TGC TAC GGC GAC ATG GAG AAC AAG CTG TTT GCG GGG

790
 ile arg arg asp gly leu leu leu arg leu val asp asp phe leu
 ATT CGG CGG GAC GGG CTG CTC CTG CGT TTG GTG GAT GAT TTC TTG

800
 810
 820
 830
 840
 850
 860
 870

FIG. 74
(CONTINUED)

102/103

```

                                880
leu val thr pro his leu thr his ala lys thr phe leu arg thr
TTG GTG ACA CCT CAC CTC ACC CAC GCG AAA ACC TTC CTC AGG ACC

                                890
leu val arg gly val pro glu tyr gly cys val val asn leu arg
CTG GTC CGA GGT GTC CCT GAG TAT GGC TGC GTG GTG AAC TTG CGG

                                910
lys thr val val asn phe pro val glu asp glu ala leu gly gly
AAG ACA GTG GTG AAC TTC CCT GTA GAA GAC GAG GCC CTG GGT GGC

                                920
thr ala phe val gln met pro ala his gly leu phe pro trp cys
ACG GCT TTT GTT CAG ATG CCG GCC CAC GGC CTA TTC CCC TGG TGC

                                930
                                940
gly leu leu leu asp thr arg thr leu glu val gln ser asp tyr
GGC CTG CTG CTG GAT ACC CGG ACC CTG GAG GTG CAG AGC GAC TAC

                                950
                                960
ser ser tyr ala arg thr ser ile arg ala ser val thr phe asn
TCC AGC TAT GCC CGG ACC TCC ATC AGA GCC AGT GTC ACC TTC AAC

                                970
arg gly phe lys ala gly arg asn met arg arg lys leu phe gly
CGC GGC TTC AAG GCT GGG AGG AAC ATG CGT CGC AAA CTC TTT GGG

                                980
                                990
val leu arg leu lys cys his ser leu phe leu asp leu gln val
GTC TTG CGG CTG AAG TGT CAC AGC CTG TTT CTG GAT TTG CAG GTG

                                1000
asn ser leu gln thr val cys thr asn ile tyr lys ile leu leu
AAC AGC CTC CAG ACG GTG TGC ACC AAC ATC TAC AAG ATC CTC CTG

                                1010
                                1020
leu gln ala tyr arg phe his ala cys val leu gln leu pro phe
CTG CAG GCG TAC AGG TTT CAC GCA TGT GTG CTG CAG CTC CCA TTT

                                1030
his gln gln val trp lys asn pro thr phe phe leu arg val ile
CAT CAG CAA GTT TGG AAG AAC CCC ACA TTT TTC CTG CGC GTC ATC

                                1040
                                1050
ser asp thr ala ser leu cys tyr ser ile leu lys ala lys asn
TCT GAC ACG GCC TCC CTC TGC TAC TCC ATC CTG AAA GCC AAG AAC

                                1060
ala gly met ser leu gly ala lys gly ala ala gly pro leu pro
GCA GGG ATG TCG CTG GGG GCC AAG GGC GCC GCC GGC CCT CTG CCC

                                1070
                                1080
ser glu ala val gln trp leu cys his gln ala phe leu leu lys
TCC GAG GCC GTG CAG TGG CTG TGC CAC CAA GCA TTC CTG CTC AAG

                                1090
leu thr arg his arg val thr tyr val pro leu leu gly ser leu
CTG ACT CGA CAC CGT GTC ACC TAC GTG CCA CTC CTG GGG TCA CTC

```

FIG. 74
(CONTINUED)

103/103

```

      1100
arg thr ala gln thr gln leu ser arg lys leu pro gly thr thr
AGG ACA GCC CAG ACG CAG CTG AGT CGG AAG CTC CCG GGG ACG ACG

      1110
leu thr ala leu glu ala ala ala asn pro ala leu pro ser asp
CTG ACT GCC CTG GAG GCC GCA GCC AAC CCG GCA CTG CCC TCA GAC

      1120
leu thr ala leu glu ala ala ala asn pro ala leu pro ser asp
CTG ACT GCC CTG GAG GCC GCA GCC AAC CCG GCA CTG CCC TCA GAC

      1130      1132
phe lys thr ile leu asp OP
TTC AAG ACC ATC CTG GAC TGA TGGCCACCCGCCACAGCCAGGCCGAGAGCAGA

CACCAGCAGCCCTGTCACGCCGGGCTCTACGTCCCAGGGAGGGAGGGGCGGCCACACCC
AGGCCCCGACCGCTGGGAGTCTGAGGCCTGAGTGAGTGTTTGGCCGAGGCCTGCATGTCC
GGCTGAAGGCTGAGTGTCCGGCTGAGGCCTGAGCGAGTGTCAGCCAAGGGCTGAGTGTC
CAGCACACCTGCCGTCTTCACTTCCCCACAGGCTGGCGCTCGGCTCCACCCAGGGCCAG
CTTTTCYTCACCAGGAGCCCGGCTTCCACTCCCCACATAGGAATAGTCCATCCCCAGATT
CGCCATTGTTACCCYTCGCCCTGCCYTCTTTGCCTTCCACCCCCACCATCCAGGTGGA
GACCCTGAGAAGGACCCTGGGAGCTCTGGGAATTTGGAGTGACCAAAGGTGTGCCCTGTA
CACAGGCGAGGACCCTGCACCTGGATGGGGGTCCCTGTGGGTCAAATTGGGGGGAGGTGC
TGTGGGAGTAAATACTGAATATATGAGTTTTTCAGTTTTTGRAAAAAAAAAAAAAAAAA
AAAAAAAAA
```

FIG. 74
(CONTINUED)

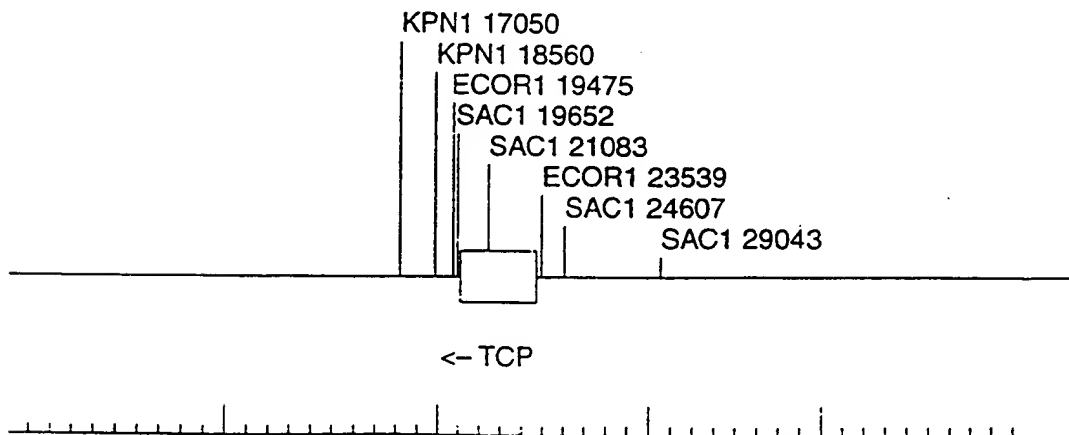


FIG. 75

4041074

5

10

10

15

20

25

1

into a senescent stage (i.e., loss of proliferative capacity). Moreover, the ability of a cell to maintain (or increase) telomere length may allow a cell to escape senescence, i.e., to become immortal.

The structure of telomeres and telomeric DNA has been investigated in numerous systems (see, e.g, Harley and Villeponteau, 1995, *Curr. Opin. Genet. Dev.* 5:249). In most organisms, telomeric DNA consists of a tandem array of very simple sequences; in humans and other vertebrates telomeric DNA consists of hundreds to thousands of tandem repeats of the sequence TTAGGG. Methods for determining and modulating telomere length in cells are described in PCT Publications WO 93/23572 and WO 96/41016.

The maintenance of telomeres is a function of a telomere-specific DNA polymerase known as *telomerase*. Telomerase is a ribonucleoprotein (RNP) that uses a portion of its RNA moiety as a template for telomere repeat DNA synthesis (Morin, 1997, *Eur. J. Cancer* 33:750; Yu et al., 1990, *Nature* 344:126; Singer and Gottschling, 1994, *Science* 266:404; Autexier and Greider, 1994, *Genes Develop.*, 8:563; Gilley et al., 1995, *Genes Develop.*, 9:2214; McEachern and Blackburn, 1995, *Nature* 367:403; Blackburn, 1992, *Ann. Rev. Biochem.*, 61:113; Greider, 1996, *Ann. Rev. Biochem.*, 65:337). The RNA components of human and other telomerases have been cloned and characterized (see, PCT Publication WO 96/01835 and Feng et al., 1995, *Science* 269:1236). However, the characterization of the protein components of telomerase has been difficult. In part, this is because it has proved difficult to purify the telomerase RNP, which is present in extremely low levels in cells in which it is expressed. For example, it has been estimated that human cells known to express high levels of telomerase activity may have only about one hundred molecules of the enzyme per cell.

Consistent with the relationship of telomeres and telomerase to the proliferative capacity of a cell (i.e., the ability of the cell to divide indefinitely), telomerase activity is detected in immortal cell lines and an extraordinarily diverse set of tumor tissues, but is not detected (i.e., was absent or below the assay threshold) in normal somatic cell cultures or normal tissues adjacent to a tumor (see, U.S. Patent Nos. 5,629,154; 5,489,508; 5,648,215; and 5,639,613; see also, Morin, 1989, *Cell*

59: 521; Shay and Bacchetti 1997, *Eur. J. Cancer* 33:787; Kim et al., 1994, *Science* 266:2011; Counter et al., 1992, *EMBO J.* 11:1921; Counter et al., 1994, *Proc. Natl. Acad. Sci. U.S.A.* 91, 2900; Counter et al., 1994, *J. Virol.* 68:3410). Moreover, a correlation between the level of telomerase activity in a tumor and the likely clinical outcome of the patient has been reported (e.g., U.S. Patent No. 5,639,613, *supra*; Langford et al., 1997, *Hum. Pathol.* 28:416). Telomerase activity has also been detected in human germ cells, proliferating stem or progenitor cells, and activated lymphocytes. In somatic stem or progenitor cells, and in activated lymphocytes, telomerase activity is typically either very low or only transiently expressed (see, Chiu et al., 1996, *Stem Cells* 14:239; Bodnar et al., 1996, *Exp. Cell Res.* 228:58; Taylor et al., 1996, *J. Invest. Dermatology* 106:759).

Human telomerase is an ideal target for diagnosing and treating human diseases relating to cellular proliferation and senescence, such as cancer. Methods for diagnosing and treating cancer and other telomerase-related diseases in humans are described in U.S. Patent Nos. 5,489,508, 5,639,613, and 5,645,986. Methods for predicting tumor progression by monitoring telomerase are described in U.S. Patent No. 5,639,613. The discovery and characterization of the catalytic protein subunit of human telomerase would provide additional useful assays for telomerase and for disease diagnosis and therapy. Moreover, cloning and determination of the primary sequence of the catalytic protein subunit would allow more effective therapies for human cancers and other diseases related to cell proliferative capacity and senescence.

BRIEF SUMMARY OF THE INVENTION

The present invention provides a recombinant polynucleotide comprising a human telomerase reverse transcriptase (hTERT) promoter sequence. For information about hTERT itself including analogues, fragments and variants and gene sequences encoding these, reference is made to parent UK Patent Application No 97 208890.4 published under Serial No 2317891. That Application claims substantially pure, synthetic or recombinant hTERT *per se* (including analogues, fragments and variants), polynucleotides encoding the same, antibodies to hTERT, transformed cells

expressing hTRT, pharmaceutical compositions comprising these materials and various methods and uses of the materials relating to medicine and pharmacology.

In preferred aspects of the invention, the promoter sequence comprises a sequence of at least 15, optionally at least 50, optionally at least at 100, optionally
5 at least 200, or optionally at least 500 nucleotides of Figure 21 or a sequence hybridizable thereto under stringent conditions.

In other preferred aspects, the promoter sequence comprises a sequence of at least 15, optionally at least 50, optionally at least at 100, optionally at least 200, or optionally at least 500 nucleotides of nucleotides 1 to 2440 of Figure 21
10 or a sequence hybridizable thereto under stringent conditions. In one particular aspect the promoter sequence preferably comprises at least the sequence of nucleotides 622 to 2440 of Figure 21 or a sequence hybridizable thereto under stringent conditions.

The aforementioned polynucleotides may further comprise at least
15 about at 15, optionally at least 50, optionally at least at 100, optionally at least 200 or optionally at least 500 nucleotides encoded by lambda phage Gφ5 (ATCC accession no. 98505).

Any of the aforementioned polynucleotides of the invention may further comprise a transcribable sequence operably linked to the hTRT promoter
20 sequence. Preferably the transcribable sequence encodes a protein other than hTRT. More preferably the transcribable sequence causes cell death.

In one preferred aspect the transcribable sequence operably linked to the hTRT promoter sequence is a gene encoding a toxin.

In another preferred aspect the hTRT promoter sequence is operably
25 linked to a gene encoding protein having an activity that is not itself toxic to a cell, but which renders the cell sensitive to an otherwise nontoxic drug. An example of such a protein which is non-toxic but sensitizes the cell is a Herpes virus thymidine kinase.

In another preferred aspect, the hTRT promoter sequence is operably
30 linked to a reporter gene. The reporter may encode a protein that is detectable by fluorescence, phosphorescence, or by virtue of possessing an enzymatic activity.

Examples of suitable detectable proteins include firefly luciferase, β -glucuronidase, β -galactosidase, chloramphenicol acetyl transferase, green fluorescent protein, enhanced green fluorescent protein, or the human secreted alkaline phosphatase.

In second aspect the present invention provides an isolated, synthetic,
5 substantially pure, or recombinant polynucleotide having a sequence that is at least about 15 nucleotides in length, optionally to at least about 100 nucleotides in length, and comprising a sequence exactly complementary or identical to a contiguous sequence of a nucleic acid encoding the hTRT promoter as set forth in Figure 21 bases 1-2440. In a preferred form the polynucleotide is an antisense oligonucleotide.

10 Furthermore, the invention provides a method of killing a cell in which an endogenous TRT is expressed, comprising introducing the polynucleotide as hereinbefore defined into the cell *in vitro*. In killing cells using this method gancyclovir administration may additionally be employed. The methods of killing cells may be used effectively to kill human cells.

15 The invention also provides a method of inhibiting expression of hTRT in a cell, comprising introducing the polynucleotide of the second aspect of the invention into the cell *in vitro*.

Another aspect of the invention is an assay for a compound that modulates hTRT promoter activity comprising contacting an hTRT promoter
20 sequence operably linked to a transcribable sequence with the compound and detecting a change in the expression or activity of the transcribable sequence expression product. In a preferred assay expression of the transcribable sequence expression product is detected. The fluorescence, phosphorescence, or an enzymatic activity of the transcribable sequence expression product may be detected.

25 The invention further provides a method of inactivating an endogenous hTRT promoter in a cell comprising introducing *in vitro* a recombinant polynucleotide capable of recombining with the endogenous hTRT promoter under conditions in which recombination occurs, wherein the recombinant polynucleotide comprises at least at about 15, optionally at least 50, optionally at least at 100,
30 optionally at least 200, or optionally at least 500 nucleotides of nucleotides 1-2440 of Figure 21.

The invention provides a polynucleotide as hereinbefore defined for use as a pharmaceutical.

The invention also provides for the use of a polynucleotide as hereinbefore defined for the manufacture of a medicament for increasing the proliferative capacity of a vertebrate cell, preferably a mammalian cell.

The invention also provides for the use of a polynucleotide as hereinbefore defined for the manufacture of a medicament for treating a condition associated with an elevated level of telomerase activity in a cell.

In another aspect, the invention provides a polynucleotide having a promoter sequence operably linked to the sequence encoding the hTERT protein. The promoter may be a promoter other than the naturally occurring hTERT promoter. In a related aspect, the invention provides an expression vector comprising the promoter of the hTERT.

The invention also provides an isolated, synthetic, substantially pure, or recombinant polynucleotide that is at least ten nucleotides in length and comprises a contiguous sequence of at least ten nucleotides that is identical or exactly complementary to a contiguous sequence in a naturally occurring hTERT gene. In some embodiments the polynucleotide is an RNA, a DNA, or contains one or more non-naturally occurring, synthetic nucleotides. In one aspect, the polynucleotide is identical or exactly complementary to the contiguous sequence of at least ten contiguous nucleotides in a naturally occurring hTERT gene. For example, the polynucleotide may be an antisense polynucleotide. In one embodiment, the antisense polynucleotide comprises at least about 20 nucleotides.

The invention also provides a cell, such as a human, mouse, or yeast cell, containing recombinant polynucleotides such as a polynucleotide with an hTERT protein coding sequence operably linked to promoter sequences as hereinbefore described. In particular aspects, the cell is a vertebrate cell, such as a cell from a mammal, for example a human, and has an increased proliferative capacity relative to a cell that is otherwise identical but does not comprise the recombinant polynucleotide or has an increased telomerase activity level relative to a cell that is otherwise identical but does not comprise the recombinant polynucleotide. In some embodiments the cell is immortal.

In related embodiments, the invention provides organisms and cells comprising a polynucleotide encoding a human telomerase reverse transcriptase polypeptide, such as a transgenic non-human organism such as a yeast, plant, bacterium, or a non-human animal, for example, a mouse. The invention also provides
5 for transgenic animals and cells from which an hTRT gene has been deleted (knocked-out) or mutated such that the gene does not express a naturally occurring hTRT gene product. Thus, in alternative embodiments, the transgenic non-human animal has a mutated telomerase gene, is an animal deficient in a telomerase activity, is an animal whose TRT deficiency is a result of a mutated gene encoding a TRT
10 having a reduced level of a telomerase activity compared to a wild-type TRT and is an animal having a mutated TRT gene with one or more mutations, including missense mutations, nonsense mutations, insertions, or deletions.

The invention also provides a method for determining whether a compound or treatment is a modulator of a telomerase reverse transcriptase activity or
15 hTRT expression. The method involves detecting or monitoring a change in activity or expression in a cell, animal or composition comprising an hTRT protein or polynucleotide following administration of the compound or treatment. In one embodiment, the method includes the steps of: providing a TRT composition, contacting the TRT with the test compound and measuring the activity of the TRT
20 where a change in TRT activity in the presence of the test compound is an indicator that the test compound modulates TRT activity. In certain embodiments, the composition is a cell, an organism, a transgenic organism or an *in vitro* system, such as an expression system, which contains a recombinant polynucleotide encoding an hTRT polypeptide. Thus, the hTRT of the method may be a product of *in vitro*
25 expression. In various embodiments the detection of telomerase activity or expression may be by detecting a change in abundance of an hTRT gene product, monitoring incorporation of a nucleotide label into a substrate for telomerase, monitoring hybridization of a probe to an extended telomerase substrate, monitoring amplification of an extended telomerase substrate, monitoring telomere length of a cell exposed to the
30 test compound, monitoring the loss of the ability of the telomerase to bind to a chromosome, or measuring the accumulation or loss of telomere structure.

In one aspect, the invention provides a method of detecting an hTERT gene product in a biological sample by contacting the biological sample with a probe that specifically binds the gene product, wherein the probe and the gene product form a complex, and detecting the complex, where the presence of the complex is correlated with the presence of the hTERT gene product in the biological sample. The gene product may be RNA, DNA or a polypeptide. Examples of probes that may be used for detection include, but are not limited to, nucleic acids and antibodies.

In one embodiment, the gene product is a nucleic acid which is detected by amplifying the gene and detecting the amplification product, where the presence of the complex or amplification product is correlated with the presence of the hTERT gene product in the biological sample.

Various methods of detecting hTERT gene products (protein or nucleic acid), methods of detecting immortal or telomerase positive cells, methods of diagnosing telomerase-related conditions or cancer, methods of prognosing cancer, or methods of monitoring the efficacy of an anticancer treatment are described and hereby referred to in parent UK Patent Application No 97208890.4 published under Serial No 2317891.

The invention also provides kits for the detection of an hTERT gene or gene product. In one embodiment, the kit includes a container including a molecule selected from an hTERT nucleic acid or subsequence thereof, an hTERT polypeptide or subsequence thereof and an anti hTERT antibody.

The invention also provides for the use of recombinant polynucleotides in methods of treating human diseases, particularly for the manufacture of medicaments for increasing the proliferative capacity of a vertebrate cell, such as a mammalian cell, wherein the recombinant polynucleotide is introduced into the cell, said polynucleotide comprising a sequence encoding an hTERT polypeptide operably linked to a promoter sequence as hereinbefore defined. In one embodiment, the hTERT polypeptide has a sequence as shown in Figure 17. In one embodiment, the hTERT has telomerase catalytic activity. In one embodiment, the cell is human, such as a cell in a human patient. In an alternative embodiment, the cell is cultured *in vitro*. In a related embodiment, the cell is introduced into a human patient.

The invention further provides a method for treating a human disease

by introducing recombinant hTRT polynucleotide as hereinbefore defined into at least one cell in a patient. In one embodiment, a gene therapy vector is used.

The invention also provides a method for treating a condition associated with an elevated level of telomerase activity within a cell, comprising
5 introducing into said cell a therapeutically effective amount of an inhibitor of said telomerase activity, wherein said inhibitor is an hTRT polynucleotide as hereinbefore defined. In one embodiment, the inhibitor is a polynucleotide comprising, e.g., at least a subsequence of a sequence shown in Figures 16, 17, or 20. In additional
10 embodiments, the polynucleotide inhibits a TRT activity, such as binding of endogenous TRT to telomerase RNA.

The invention also provides pharmacological compositions containing a pharmaceutically acceptable carrier and a molecule selected from a polynucleotide encoding an hTRT polypeptide as hereinbefore defined or an hTRT nucleic acid or subsequence thereof comprising a promoter as hereinbefore defined.

15

DESCRIPTION OF THE FIGURES

Figure 1 shows highly conserved residues in TRT motifs from human, *S. pombe* (tez1), *S. cerevisiae* (EST2) and *Euplotes aediculatus* (p123). Identical amino acids are indicated with an asterisk (*) [raised slightly], while the similar
20 amino acid residues are indicated by a dot (·). Motif "0" in the figure is also called Motif T; Motif "3" is also called Motif A.

Figure 2 shows the location of telomerase-specific and RT-specific sequence motifs of telomerase proteins and other reverse transcriptases. Locations of telomerase-specific motif T and conserved RT motifs 1, 2 and A-E are indicated by
25 boxes. The open rectangle labeled HIV-1 RT delineates the portion of this protein shown in Figure 3.

Figure 3 shows the crystal structure of the p66 subunit of HIV-1 reverse transcriptase (Brookhaven code 1HNV). The view is from the back of the right hand to enable all motifs to be shown.

30 Figure 4 shows multiple sequence alignment of telomerase RTs (Sp_Trt1p, *S. pombe* TRT [also referred to herein as "tez1p"]; hTRT, human TRT;

Ea_p123, *Euplotes* p123; Sc_Est2p, *S. cerevisiae* Est2p) and members of other RT families (Sc_al, cytochrome oxidase group II intron 1-encoded protein from *S. cerevisiae* mitochondria, Dm_TART, reverse transcriptase from *Drosophila melanogaster* TART non-LTR retrotransposable element); HIV-1, human immunodeficiency virus reverse transcriptase). TRT con and RT con represent consensus sequences for telomerase RTs and non-telomerase RTs. Amino acids are designated with an h, hydrophobic; p, polar; c, charged. Triangles show residues that are conserved among telomerase proteins but different in other RTs. The solid line below motif E highlights the primer grip region.

10 Figure 5 shows expression of hTRT RNA in telomerase-negative mortal cell strains and telomerase-positive immortal cell lines as described in Example 2.

Figure 6 shows a possible phylogenetic tree of telomerases and retroelements rooted with RNA-dependent RNA polymerases.

15 Figure 7 shows a restriction map of lambda clone Gφ5.

Figure 8 shows a map of chromosome 5p with the location of the STS marker D5S678 (located near the hTRT gene) indicated.

Figure 9 shows the construction of a hTRT promoter-reporter plasmid.

20 Figure 10, in two pages, shows coexpression *in vitro* of hTRT and hTR to produce catalytically active human telomerase.

Figure 11, in two pages, shows an alignment of sequences from four TRT protein and identifies motifs of interest. TRT con shows a TRT consensus sequence. RT con shows consensus residues for other reverse transcriptases. Consensus residues in upper case indicate absolute conservation in TRT proteins.

25 Figure 12 shows a Topoisomerase II cleavage site and NFκB binding site motifs in an hTRT intron, with the sequence shown corresponding to SEQUENCE ID NO:7.

Figure 13, in two pages, shows the sequence of the DNA encoding the *Euplotes* 123 kDa telomerase protein subunit (*Euplotes* TRT).

30 Figure 14 shows the amino acid sequence of the *Euplotes* 123 kDa telomerase protein subunit (*Euplotes* TRT protein).

Figure 15, in five pages, shows the DNA and amino acid sequences of the *S. pombe* telomerase catalytic subunit (*S. pombe* TRT).

Figure 16, in two pages, shows the hTRT cDNA sequence, with the sequence shown corresponding to SEQUENCE ID NO: 1.

5 Figure 17 shows the hTRT protein encoded by the cDNA of Figure 16. The protein sequence shown corresponds to SEQUENCE ID NO: 2.

Figure 18 shows the sequence of clone 712562, with the sequence shown corresponding to SEQUENCE ID NO: 3.

10 Figure 19 shows a 259 residue protein encoded by clone 712562, with the sequence shown corresponding to SEQUENCE ID NO: 10.

Figure 20 shows, in seven pages, the sequence of a nucleic acid with an open reading frame encoding a $\Delta 182$ variant polypeptide, with the sequence shown corresponding to SEQUENCE ID NO: 4. This Figure also shows the amino acid sequence of this $\Delta 182$ variant polypeptide, with the amino acid sequence shown
15 corresponding to SEQUENCE ID NO: 5.

Figure 21 shows, in six pages, sequence from an hTRT genomic clone, with the sequence shown corresponding to SEQUENCE ID NO: 6. Consensus motifs and elements are indicated, including sequences characteristic of a topoisomerase II cleavage site, NF κ B binding sites, an Alu sequence and other sequence elements.

20 Figure 22 shows the effect of mutation of the TRT gene in yeast, as described in Example 1.

Figure 23 shows the sequence of EST AA281296, corresponding to SEQUENCE ID NO: 8.

25 Figure 24 shows the sequence of the 182 basepairs deleted in clone 712562, with the sequence shown corresponding to SEQUENCE ID NO: 9.

Figure 25 shows the results of an assay for telomerase activity from BJ cells transfected with an expression vector encoding an hTRT protein (pGRN133) or a control plasmid (pBBS212) as described in Example 13.

30 Figure 26 is a schematic diagram of the affinity purification of telomerase showing the binding and displacement elution steps.

Figure 27 is a photograph of a Northern blot of telomerase.

preparations obtained during a purification protocol, as described in Example 1.

Lane 1 contained 1.5 fmol telomerase RNA, lane 2 contained 4.6 fmol telomerase RNA, lane 3 contained 14 fmol telomerase RNA, lane 4 contained 41 fmol telomerase RNA, lane 5 contained nuclear extract (42 fmol telomerase), lane 6 contained Affi-Gel(RTM)-heparin-purified telomerase (47 fmol telomerase), lane 7 contained affinity-purified telomerase (68 fmol), and lane 8 contained glycerol gradient-purified telomerase (35 fmol).

Figure 28 shows telomerase activity through a purification protocol.

Figure 29 is a photograph of a SDS-PAGE gel, showing the presence of an approximately 123 kDa polypeptide and an approximately 43 kDa doublet from *Euplotes aediculatus*.

Figure 30 is a graph showing the sedimentation coefficient of *Euplotes aediculatus* telomerase.

Figure 31 is a photograph of a polyacrylamide/urea gel with 36% formamide showing the substrate utilization of *Euplotes* telomerase.

Figure 32 shows the putative alignments of telomerase RNA template, and hairpin primers with telomerase RNA.

Figure 33 is a photograph of lanes 25-30 of the gel shown in Figure 31, shown at a lighter exposure level.

Figure 34 shows the DNA sequence of the gene encoding the 43 kDa telomerase protein subunit from *Euplotes*.

Figure 35 shows, in four pages, the DNA sequence, as well as the amino acid sequences of all three open reading frames of the 43 kDa telomerase protein subunit from *Euplotes*.

Figure 36 shows a sequence comparison between the 123 kDa telomerase protein subunit of *Euplotes* (upper sequence) and the 80 kDa polypeptide subunit of *T. thermophila* (lower sequence).

Figure 37 shows a sequence comparison between the 123 kDa telomerase protein subunit of *E. aediculatus* (upper sequence) and the 95 kDa telomerase polypeptide of *T. thermophila* (lower sequence).

Figure 38 shows the best-fit alignment between a portion of the "La-

domain" of the 43 kDa telomerase protein subunit of *E. aediculatus* (upper sequence) and a portion of the 95 kDa polypeptide subunit of *T. thermophila* (lower sequence).

Figure 39 shows the best-fit alignment between a portion of the "La-domain" of the 43 kDa telomerase protein subunit of *E. aediculatus* (upper sequence) and a portion of the 80 kDa polypeptide subunit of *T. thermophila* (lower sequence).

Figure 40 shows the alignment and motifs of the polymerase domain of the 123 kDa telomerase protein subunit of *E. aediculatus* and the polymerase domains of various reverse transcriptases including a cytochrome oxidase group II intron 1-encoded protein from *S. cerevisiae* mitochondria (a1 S.c. (group II)), Dong (LINE), and yeast ESTp (L8543.12).

Figure 41 shows the alignment of a domain of the 43 kDa telomerase protein subunit with various La proteins.

Figure 42 shows the nucleotide sequence encoding the *T. thermophila* 80 kDa protein subunit.

Figure 43 shows the amino acid sequence of the *T. thermophila* 80 kDa protein subunit.

Figure 44 shows the nucleotide sequence encoding the *T. thermophila* 95 kDa protein subunit.

Figure 45 shows the amino acid sequence of the *T. thermophila* 95 kDa protein subunit.

Figure 46 shows the amino acid sequence of L8543.12 ("Est2p").

Figure 47 shows the alignment of the amino acid sequence encoded by the *Oxytricha* PCR product with the *Euplotes* p123 sequence.

Figure 48 shows the DNA sequence of Est2.

Figure 49 shows partial amino acid sequence from a cDNA clone encoding human telomerase peptide motifs.

Figure 50 shows partial DNA sequence of a cDNA clone encoding human telomerase peptide motifs.

Figure 51 shows the amino acid sequence of *tez1*, also called *S. pombe trt*.

Figure 52 shows, in two pages, the DNA sequence of *tez1*. Intronic

and other non-coding regions are shown in lower case and exons (*i.e.*, coding regions) are shown in upper case.

Figure 53 shows the alignment of EST2p, *Euplotes*, and *Tetrahymena* sequences, as well as consensus sequence.

5 Figure 54 shows the sequences of peptides useful for production of anti-hTRT antibodies.

Figure 55 is a schematic summary of the *tez1*⁺ sequencing experiments.

Figure 56 shows two degenerate primers used in PCR to identify the *S. pombe* homolog of the *E. aediculatus* p123 sequences.

10 Figure 57 shows the four major bands produced in PCR using degenerate primers to identify the *S. pombe* homolog of the *E. aediculatus* p123 sequences.

Figure 58 shows the alignment of the M2 PCR product with *E. aediculatus* p123, *S. cerevisiae*, and *Oxytricha* telomerase protein sequences.

15 Figure 59 is a schematic showing the 3' RT PCR strategy for identifying the *S. pombe* homolog of the *E. aediculatus* p123.

Figure 60 shows characteristics of the libraries used to screen for *S. pombe* telomerase protein sequences and shows the results of screening the libraries for *S. pombe* telomerase protein sequences.

20 Figure 61 shows the positive results obtained with the *Hind*III-digested positive genomic clones containing *S. pombe* telomerase sequence.

Figure 62 is a schematic showing the 5' RT PCR strategy used to obtain a full length *S. pombe* TRT clone.

25 Figure 63 shows the alignment of RT domains from telomerase catalytic subunits for *S. pombe* (S.p.), *S. cerevisiae* (S.c.) and *E. aediculatus* (E.a.).

Figure 64 shows the alignment of the sequences from *Euplotes* ("Ea_p123"), *S. cerevisiae* ("Sc_Est2p"), and *S. pombe* ("Sp_Tez1p"). In Panel A, the shaded areas indicate residues shared between two sequences. In Panel B, the shaded areas indicate residues shared between all three sequences.

30 Figure 65 shows the disruption strategy used with the telomerase genes in *S. pombe*.

Figure 66 shows the experimental results confirming disruption of *tez1*.

Figure 67 shows the progressive shortening of telomeres in *S. pombe* due to *tez1* disruption.

5 Figure 68 shows, in four pages, the DNA and amino acid of the ORF encoding an approximately 63 kDa telomerase protein or fragment thereof encoded the EcoRI-NotI insert of clone 712562.

Figure 69 shows an alignment of reverse transcriptase motifs from various sources.

10 Figure 70 provides a restriction and function map of plasmid pGRN121.

Figure 71 shows, in two pages, the results of preliminary nucleic acid sequencing analysis of a hTRT cDNA sequence.

15 Figure 72 shows, in ten pages, the preliminary nucleic acid sequence of hTRT and deduced ORF sequences in three reading frames.

Figure 73 provides a restriction and function map of plasmid pGRN121.

Figure 74 shows, in eight pages, refined nucleic acid sequence and deduced ORF sequences of hTRT

20 Figure 75 shows a restriction map of lambda clone 25-1.1.

DETAILED DESCRIPTION OF THE INVENTION

I) INTRODUCTION

25 Telomerase is a ribonucleoprotein complex (RNP) comprising an RNA component and a catalytic protein component. The catalytic protein component of telomerase is hereinafter referred to as "TRT" (telomerase reverse transcriptase). TRT is so named because this protein acts as an RNA-dependent DNA polymerase (reverse transcriptase), using the telomerase RNA component (hereinafter, "TR") to direct synthesis of telomere DNA repeat sequences. Moreover, TRT is evolutionarily
30 related to other reverse transcriptases (see Example 12).

 The catalytic protein component of human telomerase is hereinafter

referred to as "hTRT." Human TRT is of extraordinary interest and value because, as noted *supra*, telomerase activity in human (and other mammalian cells) correlates with cell proliferative capacity, cell immortality, and the development of a neoplastic phenotype. For example, telomerase activity, and, as demonstrated in Example 2,
5 *infra*, levels of human TRT gene products and are elevated in immortal human cells (such as malignant tumor cells and immortal cell lines) relative to mortal cells (such as most human somatic cells).

Described herein are methods and compositions valuable for diagnosis, prognosis, and treatment of human diseases and disease conditions. Also described
10 are methods and reagents useful for immortalizing cells (*in vivo* and *ex vivo*), producing transgenic animals with desirable characteristics, and numerous other uses, many of which are described *infra*. Methods and reagents useful for preparing, cloning, or re-cloning TRT genes and proteins from ciliates, fungi, vertebrates, such as mammals, and other organisms are described.

15 As described in detail *infra*, TRT was initially characterized following purification of telomerase from the ciliate *Euplotes aediculatus*. Extensive purification of *E. aediculatus* telomerase, using RNA-affinity chromatography and other methods, yielded the protein "p123". Surprisingly, p123 is unrelated to proteins previously believed to constitute the protein subunits of the telomerase
20 holoenzyme (i.e., the p80 and p95 proteins of *Tetrahymena thermophila*). Analysis of the p123 DNA and protein sequences (Genbank Accession No. U95964; Figures 13 and 14) revealed reverse transcriptase (RT) motifs consistent with the role of p123 as the catalytic subunit of telomerase (see, e.g., Figures 1, 4 and 11). Moreover, p123 is related to a *S. cerevisiae* (yeast) protein, Est2p, which was known to play a
25 role in maintenance of telomeres in *S. cerevisiae* (Genbank Accession No. S5396), but prior to the present invention was not recognized as encoding a telomerase catalytic subunit protein (see, e.g., Lendvay et al., 1996, *Genetics*, 144:1399).

hTRT genes may be identified and cloned using nucleic acid probes and primers generated or derived from the TRT polynucleotides disclosed herein;
30 antibodies that specifically recognize the motifs or motif sequences or other TRT epitopes (e.g., for expression cloning TRT genes or purification of TRT proteins); by

screening computer databases; or other means. For example, as described in Example 1, PCR (polymerase chain reaction) amplification of *S. pombe* DNA was carried out with degenerate-sequence primers designed from the *Euplotes* p123 RT motifs B' and C. Of four prominent products generated, one encoded a peptide
5 sequence homologous to *Euplotes* p123 and *S. cerevisiae* Est2p. Using this PCR product as a probe, the complete sequence of the *S. pombe* TRT homologue was obtained by screening of *S. pombe* cDNA and genomic libraries and amplifying *S. pombe* RNA by reverse transcription and PCR (RT-PCR). The complete sequence of the *S. pombe* gene ("trt1"; GenBank Accession No. AF015783; Figure 15) revealed
10 that homology with p123 and Est2p was especially high in the reverse transcriptase motifs. *S. pombe* trt1 is also referred to as tez1.

Amplification using degenerate primers derived from the telomerase RT motifs was also used to obtain TRT gene sequences in *Oxytricha trifallax* and *Tetrahymena thermophila*, as described in Example 1.

15 The *Euplotes* p123, *S. pombe* trt1, and *S. cerevisiae* Est2p nucleic acid sequences of the invention were used in a search of a computerized database of human expressed sequence tags (ESTs) using the program BLAST (Altschul et al, 1990, *J. Mol. Biol.* 215:403). Searching this database with the Est2p sequence did not indicate a match, but searching with p123 and trt1 sequences identified a human
20 EST (Genbank accession no. AA281296; see SEQUENCE ID NO: 8), as described in Example 1, putatively encoding a homologous protein. Complete sequencing of the cDNA clone containing the EST (hereinafter, "clone 712562"; see SEQUENCE ID NO: 3) showed that seven RT motifs were present. However, this clone did not encode a contiguous human TRT with all seven motifs, because motifs B', C, D, and
25 E were contained in a different open reading frame (ORF) than the more NH₂-terminal motifs. In addition, the distance between motifs A and B' was substantially shorter than that of the three previously characterized TRTs. Clone 712562 was obtained from the I.M.A.G.E. Consortium; Lennon et al., 1996, *Genomics* 33:151.

A cDNA clone, pGRN121, encoding a functional hTRT (see Figure
30 16, SEQUENCE ID NO: 1) was isolated from a cDNA library derived from the human 293 cell line as described in Example 1. Comparing clone 712562 with

pGRN121 showed that clone 712562 has a 182 base pair (see Figure 24, SEQUENCE ID NO: 9) deletion between motifs A and B'. The additional 182 base pairs present in pGRN121 place all of the TRT motifs in a single open reading frame, and increase the spacing between the motif A and motif B' regions to a distance
5 consistent with the other known TRTs. As is described *infra* in the Examples (e.g., Example 7), SEQUENCE ID NO: 1 encodes a catalytically active telomerase protein having the sequence of SEQUENCE ID NO: 2. The polypeptide of SEQUENCE ID NO: 2 has 1132 residues and a calculated molecular weight of about 127 kilodaltons (kD).

10 As is discussed *infra*, and described in Example 9, *infra*, TRT cDNAs possessing the 182 basepair deletion characteristic of the clone 712562 are detected following reverse transcription of RNA from telomerase-positive cells (e.g., testis and 293 cells). hTRT RNAs lacking this 182 base pair sequence are referred to generally as "Δ182 variants" and may represent one, two, or several species. Although the hTRT
15 variants lacking the 182 basepair sequence found in the pGRN121 cDNA are unlikely to encode a fully active telomerase catalytic enzyme, they may play a role in telomerase regulation, as discussed *infra*, and/or have partial telomerase activity, such as telomere binding or hTR binding activity, as discussed *infra*.

Thus, in some embodiments the present invention provides an isolated
20 polynucleotide with a sequence of a naturally occurring human TRT gene. In other related embodiments the invention provides sense and antisense nucleic acids that bind to an hTRT gene. The present invention also provides many novel methods, including methods that employ the aforementioned compositions, for example, by providing
25 diagnostic and prognostic assays for human diseases, methods for developing therapeutics and methods of therapy, identification of telomerase-associated proteins, and methods for screening for agents capable of activating or inhibiting telomerase activity. Numerous other aspects and embodiments of the invention are provided *infra*.

One aspect of the invention is the use of a polynucleotide that is at least ten nucleotides to about 10 kb or more in length and comprises a contiguous
30 sequence of at least ten nucleotides that is identical or exactly complementary to a contiguous sequence in a naturally occurring hTRT gene in assaying or screening for

an hTRT gene sequence or in preparing a recombinant host cell.

A further aspect of the invention is the use of an agent increasing expression of hTRT in the manufacture of a medicament for the treatment of a condition addressed by increasing proliferative capacity of a vertebrate cell, optionally the medicament being for inhibiting the effects of aging.

Yet a further aspect of the invention is the use of an inhibitor of telomerase activity in the manufacture of a medicament for the treatment of a condition associated with an elevated level of telomerase activity within a human cell.

The polynucleotides or fragments of the invention are also each provided in a further aspect of this invention for use as a pharmaceutical.

Other embodiments of the invention further include the use of a polynucleotide or fragment, in each case as defined herein, in the manufacture of a medicament, for example in the manufacture of a medicament for inhibiting an effect of aging or cancer.

In certain embodiments of the present invention, the hTRT polynucleotides are other than the 389 nucleotide polynucleotide of SEQUENCE ID NO:8 and/or other than clone 712562, the plasmid containing an insert, the sequence of which insert is shown in Figure 18 (SEQUENCE ID NO:3).

II) TRT GENES

For more information about TRT proteins reference is hereby made to parent UK Patent Application No 97 208890.4 published under Serial No 2317891.

It will be apparent to one of skill that, provided with the reagents, including the TRT sequences disclosed herein for those reagents and the methods and guidance provided herein (including specific methodologies described *infra*), TRT genes and proteins can be obtained, isolated and produced in recombinant form by one of ordinary skill. For example, primers (e.g., degenerate amplification primers) are provided that hybridize to gene sequences encoding RT and T motifs characteristic of TRT. For example, one or more primers or degenerate primers that hybridize to sequences encoding the FFYXTE region of the T motif, other TRT motifs (as discussed *infra*), or combinations of motifs or consensus sequences, can be prepared based on the

codon usage of the target organism, and used to amplify the TRT gene sequence from genomic DNA or cDNA prepared from the target organism. Use of degenerate primers is well known in the art and entails use of sets of primers that hybridize to the set of nucleic acid sequences that can potentially encode the amino acids of the target motif, taking into account codon preferences and usage of the target organism, and by using amplification (e.g., PCR) conditions appropriate for allowing base mismatches in the annealing steps of PCR. Typically two primer sets are used; however, single primer (or, in this case, a single degenerate primer set) amplification systems are well known and may be used to obtain TRT genes.

Table 1 provides illustrative primers of the invention that may be used to amplify novel TRT nucleic acids, particularly those from vertebrates (e.g., humans and other mammals). "N" is an equimolar mixture of all four nucleotides, and nucleotides within parentheses are equimolar mixtures of the specified nucleotides.

TABLE 1
ILLUSTRATIVE DEGENERATE PRIMERS FOR AMPLIFICATION OF TRT
NUCLEIC ACIDS

	<u>motif</u>	<u>direction</u>	<u>5' - sequence -3'</u>
a	<u>FFYVTE</u>	Forward	TT(CT)TT(CT)TA(CT)GTNACNGA
b	<u>FFYVTE</u>	Reverse	TCNGTNAC(GA)TA(GA)AA(GA)AA
c	<u>RFIPKP</u>	Forward	(CA)GNTT(CT)AT(ACT)CCNAA(AG)CC
d	<u>RFIPKP</u>	Reverse	GG(TC)TTNGG(TGA)AT(GA)AANC
e	<u>AYDTI</u>	Forward	GCNTA(CT)GA(CT)ACNAT
f	<u>AYDTI</u>	Reverse	TANGT(GA)TC(GA)TANGC
g	<u>GIPQG</u>	Forward	GGNAT(ACT)CCNCA(AG)GG
h	<u>GIPQGS</u>	Reverse	(GC)(AT)NCC(TC)TGNGG(TGA)ATNCC
i	<u>LVDDFL</u>	Forward	(CT)TNGTNGA(CT)GA(CT)TT(CT)(CT)T
j	<u>DDFLLVT</u>	Reverse	GTNACNA(GA)NA(GA)(GA)AA(GA)TC(GA)TC

Preferred primer combinations (y = yes, n = no)

		<u>Reverse</u>				
<u>Forward</u>		<u>b</u>	<u>d</u>	<u>f</u>	<u>h</u>	<u>i</u>
	a -	n	y	y	y	y
5	c -	n	n	y	y	y
	e -	n	n	n	y	y
	g -	n	n	n	n	y
	i -	n	n	n	n	n

10 Amplified TRT nucleic acid may be used as a hybridization probe for colony hybridization to a library (e.g., cDNA library) made from the target organism, such that a nucleic acid having the entire TRT protein coding sequence, or a substantial portion thereof, is identified and isolated or cloned. Reagents and methods such as those just described were used in accordance with the methods described herein to
15 obtain TRT gene sequences of *Oxytricha trifallax* and *Tetrahymena thermophila*, as described in detail *infra*. It will be recognized that following cloning of a previously uncharacterized TRT gene, the sequence can be determined by routine methods and the encoded polypeptide synthesized and assayed for a TRT activity, such as telomerase catalytic activity (as described herein and/or by telomerase assays known in the art).

20 It will also be apparent to those of skill that TRT genes may be cloned using any of a variety of cloning methods because the TRT motif sequences and the nucleic acids of the invention comprising such sequences can be used in a wide variety of such methods. For example, hybridization using a probe based on the sequence of a known TRT to DNA or other nucleic acid libraries from the target organism, as
25 described in Example 1 can be used. It will be appreciated that degenerate PCR primers or their amplification products such as those described *supra*, may themselves be labeled and used as hybridization probes. Expression cloning methods may be used. For example, one or more antibodies that specifically bind peptides that span a TRT motif or other TRT epitope, such as the FFYXTE motif can be employed to
30 isolate a ribosomal complex comprising a TRT protein and the mRNA that encodes it. For generating such antibodies peptide immunogens are used typically between 6 and 30 amino acids in length, more often about 10 to 20 amino acids in length. The antibodies

may also be used to probe a cDNA expression library derived from the organism of interest to identify a clone encoding a TRT sequence. Computer searches of DNA databases for DNAs containing sequences conserved with known TRTs can also be used to identify a clone comprising TRT sequence.

5 Usually the naturally occurring TRT has a molecular weight of between about 80,000 daltons (D) and about 150,000 D, most often between about 95,000 D and about 130,000 D. Typically, the naturally occurring TRT has a net positive charge at pH 7 (calculated pI typically greater than 9). The present invention also encompasses isolated or recombinant polynucleotide having the sequence of a naturally occurring
10 gene encoding a TRT protein. The invention may therefore provide reagents useful for isolating sequence of a TRT from nonvertebrate (such as a yeast) and vertebrates, such as mammals (e.g., murine or human). The isolated polynucleotide may be associated with other naturally occurring or recombinant or synthetic vector nucleic acid sequences. Typically, the isolated nucleic acid is smaller than about 300 kb, often less
15 than about 50 kb, more often less than about 20 kb, frequently less than about 10 kb and sometimes less than about 5 kb or 2 kb in length. In some embodiments the isolated TRT polynucleotide is even smaller, such as a gene fragment, primer, or probe of less than about 1 kb or less than 0.1 kb.

20 III) NUCLEIC ACIDS

A) GENERALLY

Isolated and recombinant nucleic acids having a sequence of a polynucleotide encoding a telomerase catalytic subunit protein (TRT), such as a recombinant TRT gene from *Euplotes*, *Tetrahymena*, *S. pombe* or humans are shown in
25 Figure 13 (*Euplotes*); Figure 15 (*S. pombe*) and Figure 16 (human, GenBank Accession No. AF015950). The present invention encompasses sense and anti-sense polynucleotides having a TRT gene sequence, including probes, primers, TRT-protein-encoding polynucleotides, and the like.

30 B) HUMAN TRT

The present invention encompasses nucleic acids having a sequence of a

telomerase catalytic subunit from humans (i.e., hTRT).

In one aspect, the invention encompasses a polynucleotide having a sequence or subsequence of a human TRT gene. The polynucleotide may have a sequence of SEQUENCE ID NO: 3 (Figure 18), SEQUENCE ID NO: 4 (Figure 20),
5 or subsequences thereof. The invention also provides polynucleotides with substantial sequence identity to the hTRT nucleic acid sequences disclosed herein, e.g., including but not limited to SEQUENCE ID NOS: 4 [Figure 20] and 6 [Figure 21], and 7. The invention also encompasses naturally occurring alleles of human TRT genes and variant polynucleotide sequences having one or more nucleotide deletions, insertions or
10 substitutions relative to an hTRT nucleic acid sequence disclosed herein. As described *infra*, variant nucleic acids may be produced using the recombinant or synthetic methods described below or by other means.

The invention thus provides isolated and recombinant polynucleotides having a sequence from a flanking region of a human TRT gene. Such polynucleotides
15 include those derived from genomic sequences of untranslated regions of the hTRT mRNA. An exemplary genomic sequence is shown in Figure 21 (SEQUENCE ID NO: 6). As described in Example 4, SEQUENCE ID NO: 6 was obtained by sequencing a clone, λ G Φ 5 isolated from a human genomic library. Lambda G Φ 5 contains a 15 kilobasepair (kbp) insert including approximately 13,000 bases 5' to the hTRT coding
20 sequences. This clone contains hTRT promoter sequences and other hTRT gene regulatory sequences (e.g., enhancers).

The invention also encompasses isolated and recombinant polynucleotides having a sequence from an intronic region of a human TRT gene. An exemplary intronic sequence is shown in Figure 21 (SEQUENCE ID NO: 7; see
25 Example 3). In some embodiments, hTRT introns are included in "minigenes" for improved expression of hTRT proteins in eukaryotic cells.

In a related aspect, the present invention encompasses polynucleotides that encode hTRT proteins or protein fragments, including modified, altered and variant hTRT polypeptides. In one embodiment, the encoded hTRT protein or fragment has an
30 amino acid sequence as set forth in Figure 17 (SEQUENCE ID NO: 2), or with conservative substitutions of SEQUENCE ID NO: 2. In one embodiment, the encoded

hTERT protein or fragment has substitutions that change an activity of the protein (e.g., telomerase catalytic activity).

It will be appreciated that, as a result of the degeneracy of the genetic code, the nucleic acid encoding the hTERT protein need not have the sequence of a naturally occurring hTERT gene, but that a multitude of polynucleotides can encode an hTERT polypeptide having an amino acid sequence of SEQUENCE ID NO: 2. The present invention provides each and every possible variation of nucleotide sequence that could be made by selecting combinations based on possible codon choices made in accordance with known triplet genetic codes, and all such variations are specifically disclosed hereby. Thus, although in some cases hTERT polypeptide-encoding nucleotide sequences that are capable of hybridizing to the nucleotide sequence of the naturally occurring sequence (under appropriately selected conditions of stringency) are preferred, it may be advantageous in other cases to produce nucleotide sequences encoding hTERT that employ a substantially different codon usage and so perhaps do not hybridize to nucleic acids with the naturally occurring sequence.

In particular embodiments, the invention provides hTERT oligo- and polynucleotides that comprise a subsequence of an hTERT nucleic acid disclosed herein (e.g., SEQUENCE ID NOS: 1 and 6). The nucleic acids of the invention typically comprise at least about 10, more often at least about 12 or about 15 consecutive bases of the exemplified hTERT polynucleotide. Often, the nucleic acid of the invention will comprise a longer sequence, such as at least about 25, about 50, about 100, about 200, or at least about 500 to 3000 bases in length, for example when expression of a polypeptide, or full length hTERT protein is intended.

In still other embodiments, the present invention provides "Δ182 hTERT" polynucleotides having a sequence identical or complementary to naturally occurring or non-naturally occurring hTERT polynucleotides such as SEQUENCE ID NO: 3 or SEQUENCE ID NO: 4, which do not contain the 182 nucleotide sequence (SEQUENCE ID NO: 9 [Figure 24]) found in pGRN121 (and also absent in clone 712562). These polynucleotides are of interest, in part, because they encode polypeptides that contain different combinations or arrangements of TRT motifs than found in the "full-length" hTERT polypeptide (SEQUENCE ID NO: 2) such as is

encoded by pGRN121. As discussed *infra*, it is contemplated that these polypeptides may play a biological role in nature (e.g., in regulation of telomerase expression in cells) and/or find use as therapeutics (e.g., as dominant-negative products that inhibit function of wild-type proteins), or have other roles and uses, e.g. as described herein.

5 For example, in contrast to the polypeptide encoded by pGRN121, clone 712562 encodes a 259 residue protein with a calculated molecular weight of approximately 30 kD (hereinafter, "712562 hTRT"). The 712562 hTRT polypeptide (SEQUENCE ID NO: 10 [Figure 19]) contains motifs T, 1, 2, and A, but not motifs B', C, D and E (See Figure 4). Similarly, a variant hTRT polypeptide with therapeutic and
10 other activities may be expressed from a nucleic acid similar to the pGRN121 cDNA but lacking the 182 basepairs missing in clone 712562, e.g., having the sequence shown in Figure 20 (SEQUENCE ID NO: 4). This nucleic acid (hereinafter, "pro90 hTRT"), which may be synthesized using routine synthetic or recombinant methods as described herein, encodes a protein of 807 residues (calculated molecular weight of approximately
15 90 kD) that shares the same amino terminal sequence as the hTRT protein encoded by SEQUENCE ID NO: 1, but diverges at the carboxy-terminal region (the first 763 residues are common, the last 44 residues of pro90 hTRT are different than "full-length" hTRT). The pro90 hTRT polypeptide contains motifs T, 1, 2, and A, but not motifs B, C, D, E, and thus may have some, but not likely all telomerase activities.

20

C) PRODUCTION OF HUMAN TRT NUCLEIC ACIDS

The polynucleotides of the invention have numerous uses including, but not limited to, expression of polypeptides encoding hTRT or fragments thereof, use as sense or antisense probes or primers for hybridization and/or amplification of naturally
25 occurring hTRT genes (e.g. for diagnostic or prognostic applications), and as therapeutic agents (e.g., in antisense, triplex, or ribozyme compositions). As will be apparent upon review of the disclosure, these uses will have enormous impact on the diagnosis and treatment of human diseases relating to aging, cancer, and fertility as well as the growth, reproduction, and manufacture of cell-based products. As described in
30 the following sections, the hTRT nucleic acids of the invention may be made (e.g., cloned, synthesized, or amplified) using techniques well known in the art.

1) CLONING, AMPLIFICATION, AND RECOMBINANT PRODUCTION

hTERT genes are cloned using a nucleic acid probe that specifically
5 hybridizes to hTERT genomic DNA. One suitable probe for this purpose is a
polynucleotide having all or part of the sequence provided in Figure 16 (SEQUENCE
ID NO: 1), such as a probe comprising a subsequence thereof. Typically, the target
hTERT genomic DNA is ligated into a vector (e.g., a plasmid, phage, virus, yeast artificial
chromosome, or the like) and may be isolated from a genomic library. Once an hTERT
10 nucleic acid is identified, it can be isolated according to standard methods known to
those of skill in the art. An illustrative example but not of the invention is the screening
of a human cDNA library for the hTERT gene as provided in Example 1; similarly, an
example of screening a human genomic library is found in Examples 3 and 4. Cloning
methods are well known and are described, for example, in Sambrook et al., (1989)
15 MOLECULAR CLONING: A LABORATORY MANUAL, 2ND ED., VOLS. 1-3, Cold Spring
Harbor Laboratory hereinafter, "Sambrook"); Berger and Kimmel, (1987) METHODS IN
ENZYMOLOGY, VOL. 152: GUIDE TO MOLECULAR CLONING TECHNIQUES, San Diego:
Academic Press, Inc.; Ausubel et al., CURRENT PROTOCOLS IN MOLECULAR BIOLOGY,
Greene Publishing and Wiley-Interscience, New York (1997); Cashion et al., U.S.
20 Patent No. 5,017,478; and Carr, European Patent No. 0,246,864.

The invention also provides hTERT genomic nucleic acids isolated by
amplification methods such as the polymerase chain reaction (PCR). For example, but
not of the invention, hTERT protein coding sequence is amplified from an RNA or cDNA
sample (e.g., double stranded placental cDNA (Clontech, Palo Alto CA)) using the
25 primers 5'-GTGAAGGCACTGTTTCAGCG-3' ("TCP1.1") and
5'-CGCGTGGGTGAGGTGAGGTG-3' ("TCP 1.15"). In some embodiments a third
primer or second pair of primers may be used, e.g., for "nested PCR", to increase
specificity. One example of a second pair of primers is 5'-
CTGTGCTGGGCCTGGACGATA-3' ("billTCP6") and 5'-
30 AGCTTGTTCTCCATGTCGCCGTAG-3' ("TCP1.14"). It will be apparent to those of
skill that numerous other primers and primer combinations, useful for amplification of

hTRT nucleic acids can be provided.

The invention thus provides primers that amplify promoter regions of hTRT genomic DNA. For example, the hTRT intron at position 274/275 of SEQUENCE ID NO: 1 (see Example 3) may be amplified (e.g., for detection of genomic clones) using primers TCP1.57 and TCP1.52 (primer pair 1) or primers TCP1.49 and TCP1.50 (primer pair 2). (Primer names refer to primers listed in Table 2, *infra*.) The primer pairs can be used individually or in a nested PCR where primer set 1 is used first.

Another comparative example of primers concerns those that specifically amplify and so detect the 5' end of the hTRT mRNA or the exon encoding the 5' end of hTRT gene (e.g., to assess the size or completeness of a cDNA clone). The following primer pairs are useful for amplifying the 5' end of hTRT: primers K320 and K321 (primer pair 3); primers K320 and TCP1.61 (primer pair 4); primers K320 and K322 (primer pair 5). The primer sets can be used in a nested PCR in the order set 5, then set 4 or set 3, or set 4 or set 5, then set 3. Yet another comparative example involves primers chosen to amplify or detect specifically the conserved hTRT TRT motif region comprising approximately the middle third of the mRNA (e.g., for use as a hybridization probe to identify TRT clones from, for example, nonhuman organisms). The following primer pairs are useful for amplifying the TRT motif region of hTRT nucleic acids: primers K304 and TCP1.8 (primer pair 6), or primers LT1 and TCP1.15 (primer pair 7). The primer sets can be used in a nested PCR experiment in the order set 6 then set 7.

Suitable PCR amplification conditions are known to those of skill and include (but are not limited to) 1 unit Taq polymerase (Perkin Elmer, Norwalk CT), 100 μ M each dNTP (dATP, dCTP, dGTP, dTTP), 1x PCR buffer (50 mM KCl, 10 mM Tris, pH 8.3 at room temperature, 1.5 mM MgCl₂, 0.01% gelatin) and 0.5 μ M primers, with the amplification run for about 30 cycles at 94° for 45 sec, 55° for 45 sec and 72° for 90 sec. It will be recognized by those of skill in the art that other thermostable DNA polymerases, reaction conditions, and cycling parameters will also provide suitable amplification. Other suitable *in vitro* amplification methods that can be used to obtain hTRT nucleic acids include, but are not limited to, those herein, *infra*. Once amplified,

the hTERT nucleic acids can be cloned, if desired, into any of a variety of vectors using routine molecular biological methods or detected or otherwise utilized in accordance with the methods of the invention.

One of skill will appreciate that the cloned or amplified hTERT nucleic acids obtained as described above can be prepared or propagated using other methods, such as chemical synthesis or replication by transformation into bacterial systems, such as *E. coli* (see, e.g., Ausubel et al., *supra*), or eukaryotic, such as mammalian, expression systems. Similarly, hTERT RNA can be expressed in accordance with the present *in vitro* methods, or in bacterial systems such as *E. coli* using, for example, commercially available vectors containing promoters recognized by an RNA polymerase such as T7, T3 or SP6, or transcription of DNA generated by PCR amplification using primers containing an RNA polymerase promoter.

The present invention further provides altered or modified hTERT nucleic acids. It will be recognized by one of skill that the cloned or amplified hTERT nucleic acids obtained can be modified (e.g., truncated, derivatized, altered) by methods well known in the art (e.g., site-directed mutagenesis, linker scanning mutagenesis) or simply synthesized de novo as described below. The altered or modified hTERT nucleic acids are useful for a variety of applications, including, but not limited to, facilitating cloning or manipulation of an hTERT gene or gene product, or expressing a variant hTERT gene product. For example, in one embodiment, the hTERT gene sequence is altered such that it encodes an hTERT polypeptide with altered properties or activities, as discussed in detail in *infra*, for example, by mutation in a conserved motif of hTERT. In another illustrative example, the mutations in the protein coding region of an hTERT nucleic acid may be introduced to alter glycosylation patterns, to change codon preference, to produce splice variants, remove protease-sensitive sites, create antigenic domains, modify specific activity, and the like. In other embodiments, the nucleotide sequence encoding hTERT and its derivatives is changed without altering the encoded amino acid sequences, for example, the production of RNA transcripts having more desirable properties, such as increased translation efficiency or a greater or a shorter half-life, compared to transcripts produced from the naturally occurring sequence. In yet another embodiment, altered codons are selected to increase the rate at which expression of the

peptide occurs in a particular prokaryotic or eukaryotic expression host in accordance with the frequency with which particular codons are utilized by the host. Useful *in vitro* and *in vivo* recombinant techniques that can be used to prepare variant hTRT polynucleotides of the invention are found in Sambrook et al. and Ausubel et al., both *supra*.

As noted *supra*, the present invention provides nucleic acids having flanking (5' or 3') and intronic sequences of the hTRT gene. The nucleic acids are of interest, *inter alia*, because they contain promoter and other regulatory elements involved in hTRT regulation and useful for expression of hTRT and other recombinant proteins or RNA gene products. It will be apparent that, in addition to the nucleic acid sequences provided in SEQUENCE ID NOS: 6 and 7, additional hTRT intron and flanking sequences may be readily obtained using routine molecular biological techniques. For example, additional hTRT genomic sequence may be obtained from Lambda clone GΦ5 (ATCC Accession No. 209024), described *supra* and in Example 4. Still other hTRT genomic clones and sequences may be obtained by screening a human genomic library using an hTRT nucleic acid probe having a sequence or subsequence from SEQUENCE ID NO: 1. Additional clones and sequences (e.g., still further upstream) may be obtained by using labeled sequences or subclones derived from λGΦ5 to probe appropriate libraries. Other useful methods for further characterization of hTRT flanking sequences include those general methods described by Gobinda et al., 1993, *PCR Meth. Applic.* 2:318; Triglia et al., 1988, *Nucleic Acids Res.* 16:8186; Lagerstrom et al., 1991, *PCR Methods Applic.* 1:111; and Parker et al., 1991, *Nucleic Acids Res.* 19:3055.

Intronic sequences can be identified by routine means such as by comparing the hTRT genomic sequence with hTRT cDNA sequences (see, e.g., Example 3), by S1 analysis (see Ausubel et al., *supra*, at Chapter 4), or various other means known in the art. Intronic sequences can also be found in pre-mRNA (i.e., unspliced or incompletely spliced mRNA precursors), which may be amplified or cloned following reverse transcription of cellular RNA.

When desired, the sequence of the cloned, amplified, or otherwise synthesized hTRT or other TRT nucleic acid can be determined or verified using DNA

sequencing methods well known in the art (see, e.g., Ausubel et al., *supra*). Useful methods of sequencing employ such enzymes as the Klenow fragment of DNA polymerase I, Sequenase (RTM)(US Biochemical Corp, Cleveland OH), *Taq* DNA polymerase (Perkin Elmer, Norwalk CT), thermostable T7 polymerase (Amersham, Chicago IL), or combinations of recombinant polymerases and proofreading exonucleases such as the ELONGASE Amplification System marketed by Gibco BRL (Gaithersburg MD). When sequencing or verifying the sequence of oligonucleotides (such as oligonucleotide made *de novo* by chemical synthesis), the method of Maxam and Gilbert may be preferred (Maxam and Gilbert, 1980, *Meth. Enz.* 65:499; Ausubel et al., *supra*, Ch. 7).

The 5' untranslated sequences of hTERT or other TERT mRNAs can be determined directly by cloning a "full-length" hTERT or other cDNA using standard methods such as reverse transcription of mRNA, followed by cloning and sequencing the resulting cDNA. Preferred oligo(dT)-primed libraries for screening or amplifying full length cDNAs that have been size-selected to include larger cDNAs may be preferred. Random primed libraries are also suitable and often include a larger proportion of clones that contain the 5' regions of genes. Other well known methods for obtaining 5' RNA sequences, such as the RACE protocol described by Frohman et al., 1988, *Proc. Nat. Acad. Sci USA* 85:8998, may also be used. If desired, the transcription start site of an hTERT or other TERT mRNA can be determined by routine methods using the nucleic acids provided herein (e.g., having a sequence of SEQUENCE ID NO: 1). One method is S1 nuclease analysis (Ausubel et al., *supra*) using a labeled DNA having a sequence from the 5' region of SEQUENCE ID NO: 1.

25 2) CHEMICAL SYNTHESIS OF NUCLEIC ACIDS

The present invention also provides hTRT polynucleotides (RNA, DNA or modified) that are produced by direct chemical synthesis. Chemical synthesis is generally preferred for the production of oligonucleotides or for oligonucleotides and polynucleotides containing nonstandard nucleotides (e.g., probes, primers and antisense oligonucleotides). Direct chemical synthesis of nucleic acids can be accomplished by methods known in the art, such as the phosphotriester method of Narang et al., 1979,

Meth. Enzymol. 68:90; the phosphodiester method of Brown et al., *Meth. Enzymol.* 68:109 (1979); the diethylphosphoramidite method of Beaucage et al., *Tetra. Lett.*, 22:1859 (1981); and the solid support method of U.S. Patent No. 4,458,066. Chemical synthesis typically produces a single stranded oligonucleotide, which may be converted
5 into double stranded DNA by hybridization with a complementary sequence, or by polymerization with a DNA polymerase and an oligonucleotide primer using the single strand as a template. One of skill will recognize that while chemical synthesis of DNA is often limited to sequences of about 100 or 150 bases, longer sequences may be obtained by the ligation of shorter sequences or by more elaborate synthetic methods.

10 It will be appreciated that the polynucleotides and oligonucleotides of the invention can be made using nonstandard bases (e.g., other than adenine, cytidine, guanine, thymine, and uridine) or nonstandard backbone structures to provides desirable properties (e.g., increased nuclease-resistance, tighter-binding, stability or a desired T_M).

Techniques for rendering oligonucleotides nuclease-resistant include those described in
15 PCT publication WO 94/12633. A wide variety of useful modified oligonucleotides may be produced, including oligonucleotides having a peptide-nucleic acid (PNA) backbone (Nielsen et al., 1991, *Science* 254:1497) or incorporating 2'-O-methyl ribonucleotides, phosphorothioate nucleotides, methyl phosphonate nucleotides, phosphotriester nucleotides, phosphorothioate nucleotides, phosphoramidates. Still other useful
20 oligonucleotides may contain alkyl and halogen-substituted sugar moieties comprising one of the following at the 2' position: OH, SH, SCH₃, F, OCN, OCH₃OCH₃, OCH₃O(CH₂)_nCH₃, O(CH₂)_nNH₂ or O(CH₂)_nCH₃ where n is from 1 to about 10; C₁ to C₁₀ lower alkyl, substituted lower alkyl, alkaryl or aralkyl; Cl; Br; CN; CF₃; OCF₃; O-, S-, or N-alkyl; O-, S-, or N-alkenyl; SOCH₃; SO₂CH₃; ONO₂; NO₂; N₃; NH₂;
25 heterocycloalkyl; heterocycloalkaryl; aminoalkylamino; polyalkylamino; substituted silyl; an RNA cleaving group; a cholesteryl group; a folate group; a reporter group; an intercalator; a group for improving the pharmacokinetic properties of an oligonucleotide; or a group for improving the pharmacodynamic properties of an oligonucleotide and other substituents having similar properties. Folate, cholesterol or
30 other groups which facilitate oligonucleotide uptake, such as lipid analogs, may be conjugated directly or via a linker at the 2' position of any nucleoside or at the 3' or 5'

position of the 3'-terminal or 5'-terminal nucleoside, respectively. One or more such conjugates may be used. Oligonucleotides may also have sugar mimetics such as cyclobutyls in place of the pentofuranosyl group. Other embodiments may include at least one modified base form or "universal base" such as inosine, or inclusion of other nonstandard bases such as queosine and wybutosine as well as acetyl-, methyl-, thio- and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases. The invention further provides oligonucleotides having backbone analogues such as phosphodiester, phosphorothioate, phosphorodithioate, methylphosphonate, phosphoramidate, alkyl phosphotriester, sulfamate, 3'-thioacetal, methylene(methylimino), 3'-N-carbamate, morpholino carbamate, chiral-methyl phosphonates, nucleotides with short chain alkyl or cycloalkyl intersugar linkages, short chain heteroatomic or heterocyclic intersugar ("backbone") linkages, or $\text{CH}_2\text{-NH-O-CH}_2$, $\text{CH}_2\text{-N(CH}_3\text{)-OCH}_2$, $\text{CH}_2\text{-O-N(CH}_3\text{)-CH}_2$, $\text{CH}_2\text{-N(CH}_3\text{)-N(CH}_3\text{)-CH}_2$ and $\text{O-N(CH}_3\text{)-CH}_2\text{-CH}_2$ backbones (where phosphodiester is O-P-O-CH_2), or mixtures of the same. Also useful are oligonucleotides having morpholino backbone structures (U.S. Patent No. 5,034,506).

Useful references include Oligonucleotides and Analogues, A Practical Approach, edited by F. Eckstein, IRL Press at Oxford University Press (1991); Antisense Strategies, Annals of the New York Academy of Sciences, Volume 600, Eds. Baserga and Denhardt (NYAS 1992); Milligan et al., 9 July 1993, J. Med. Chem. 36(14):1923-1937; Antisense Research and Applications (1993, CRC Press), in its entirety and specifically Chapter 15, by Sanghvi, entitled "Heterocyclic base modifications in nucleic acids and their applications in antisense oligonucleotides." Antisense Therapeutics, ed. Sudhir Agrawal (Humana Press, Totowa, New Jersey, 1996).

D) LABELING NUCLEIC ACIDS

It is often useful to label the nucleic acids of the invention, for example, when the oligonucleotides or polynucleotides are to be used as nucleic acid probes. The labels (*see infra*) may be incorporated by any of a number of means well known to those

of skill in the art. In one embodiment, an unamplified nucleic acid is labeled. Means of producing labeled nucleic acids are well known to those of skill in the art and include, for example, nick-translation, random primer labeling, end-labeling (e.g. using a kinase), and chemical conjugation (e.g., photobiotinylation) or synthesis. In another
5 embodiment, the label is simultaneously incorporated during an amplification step in the preparation of the sample nucleic acids. Thus, for example, polymerase chain reaction (PCR) or other nucleic acid amplification method with labeled primers or labeled nucleotides will provide a labeled amplification product. In another embodiment, transcription amplification using a labeled nucleotide (e.g. fluorescein-labeled UTP
10 and/or CTP) incorporates a label into the transcribed nucleic acids. An amplification product may also, or alternatively, be labeled after the amplification is completed.

E) ILLUSTRATIVE OLIGONUCLEOTIDES

As noted *supra* and discussed in detail *infra*, oligonucleotides are used
15 for a variety of uses including as primers, probes, therapeutic or other antisense oligonucleotides, triplex oligonucleotides, and numerous other uses as apparent from this disclosure. Table 2 provides certain illustrative specific oligonucleotides. It will be appreciated that numerous other useful oligonucleotides may be synthesized by one of skill, following the guidance provided herein.

20 In Table 2, "seq" means that the primer has been used, or is useful, for sequencing; "PCR" means that the primer has been used, or is useful, for PCR; "AS" means that means that the primer has been used, or is useful for antisense inhibition of telomerase activity; "CL" means that the primer has been used, or is useful in cloning regions of hTERT genes, "mut" means that the primer has been used, or is useful for
25 constructing mutants of hTERT genes. "UC" means "upper case," and "lc" means "lower case." Mismatches and insertions (relative to SEQUENCE ID NO: 1) are indicated by underlining; deletions are indicated by a "-". It will be appreciated that nothing in Table 2 is intended to limit the use of any particular oligonucleotide to any single use or set of uses.

TABLE 2
USEFUL OLIGONUCLEOTIDES

primer	5'-sequence-3'	Nuc3	mismatch ²	USE				
				seq.	PCR	AS	CL	MUT
TCP1.1	GTGAAGGCACCTGTTCAAGC			x	x			
TCP1.2	GTGATGATTTCTGTTGG			x	x			
TCP1.4	CTGGACACTCAGCCCTTGG			x	x			
TCP1.5	GCGAGGTGTCGCGACCT			x	x			
TCP1.6	TTTGATGATGCTGGCGATG			x	x			
TCP1.7	GCGGCTCGTCTTCTACAGG			x	x			
TCP1.8	CAGCAGGAGGATCTTGTAG			x	x			
TCP1.9	TGACCCCAAGGAGTGGCAG			x	x			
TCP1.10	TCAGCTGACTCGACACCG			x	x			
TCP1.11	CGGCGTGACAGGGCTGC			x	x			
TCP1.12	GCTGAAGGCTGAGTGTCC			x	x			
TCP1.13	TAGTCCATGTTACAAATCG			x	x			
TCP1.14	CTGTGCTGGGCTGGACGATA			x	x			
TCP1.15	CGCGTGGGTGAGGTGAGGTG			x	x			
TCP1.16	TTTCGGTGTGAGTGTTC			x	x			
TCP1.17	GTCACCGTGTGGGCAAG			x	x			
TCP1.19	GCTACCTGCCCAACACGG			x	x			
TCP1.20	GCGCGAAGAACGTGCTGG			x	x			
TCP1.21	CA-CTGCTCCCTGTGCGCTG		Y	x	x			
TCP1.22	TTCCCAAGGACTTGTTC			x	x			
TCP1.24	TGTTCTCAAGACGCACTG		Y	x	x			
TCP1.25	TACTGGGTGGTGGTATG			x	x			
TCP1.26	GGTCTTGGGCTGAAAGTGT			x	x			
TCP1.27	TGGTTCACCTGCTGGCAG			x	x			
TCP1.28	GTGGTTTCTGTGTGTGTC			x	x			
TCP1.29	GACACACACAGAAACAC			x	x			
TCP1.30	GTGCCAAGCAGTGAACAG			x	x			
TCP1.32B	GCAGTGCGTCTTGAGGAGC			x	x			
TCP1.33	TGGAAACCATAGCGTCAGGGAG			x	x			
TCP1.34	GCCCTCCCTGACGCTATGTT			x	x			
TCP1.35	GC(GT)CGGGCGCTGCCACTCAGG			x	x			
TCP1.35i	GCTCGGCGCTGCCACTCAGG			x				
TCP1.36	ACGCCGAGAACCAAGCACTTC			x	x			

TABLE 2
(cont.)

TCP1.38	CCAAAGAGAGTGGCTTCTTCG	x	x
TCP1.39	AAGGCAGACGTTCTTCGC	x	x
TCP1.40	CACGTTGCGGGGCGCTTG	x	x
TCP1.41	CCCTCACCAACGAGCGTCCG	x	x
TCP1.42	GGCGACGACGTGCTGGTTC	x	x
TCP1.43	GGCTCAGGGGACGGCCAC	x	x
TCP1.44	CTGGCAGGTGTACGGCTTC	x	x
TCP1.45	GGGTGGAACCGAGTGACCGTGGTTTC	x	x
TCP1.46	GACGTGGTGGCCGCGATGTGG	x	x
TCP1.47	GAACTCTGGCCGTTCGCCAAGAG	x	x
TCP1.48	GACACCAACAGAAACCAAGGTCAC	x	x
TCP1.49	CGCCCCCTCTTCGGCCAGGT	x	x
TCP1.50	CGAAGCCGAAAGGCCAGCACGTTCTT	x	x
TCP1.51	GGTGGCCCGAGTGTGCAAGAG	x	x
TCP1.52	GTAGCTGGCGACGCTGGTGTGAAG	x	x
TCP1.53	TGGCGACGACGTTGCTGTTCA	x	x
TCP1.54	TATGGTTCAGGGCCGTTCCGATTC	x	x
TCP1.55	CCAGCTGGCCCTACCAGGTGTGC	x	x
TCP1.56	GGCCCTCCCTGACGCTATGGTTCAG	x	x
TCP1.57	GGTGTGCCGCTGGCCACGTTCCG	x	x
TCP1.58	TCCCAAGGGACGACACCAAGGCACT	x	x
TCP1.59	GTACAGGGACACACTTTGTGACTTC	x	x
TCP1.60	TCGACGACGTACACACTCATCAAGCC	x	x
TCP1.61	AGCGGCAGCACTCGCGGTAGTGGC	x	x
TCP1.62	CCACCAAGCTCCTTCAGGCAAGACAC	x	x
TCP1.63	CCAGGGCTTCCACAGTGGCGAAGCAG	x	x
TCP1.64	CGCAGGAACGTGGCCAGCGGCAACA	x	x
TCP1.65	TGACCGTGGTTTCTGTGTGTGT	x	x
TCP1.66	CCCTCTTCAMGTGCTGTCTGATTCC	x	x
TCP1.67	ATCGCGGCCACCAAGTCCCT	x	x
TCP1.68	TGCTCCAGACACTCGGCCGGTAGAA	x	x
TCP1.69	ACGAAGCCGTACACCTGCC	x	x
TCP1.72	CGACATCCCTGGCGTTCTTGGCTTTC	x	x
TCP1.73	CACGTGCTGGCTCATTTCAAGGG	x	x
TCP1.74	GCGACATGGAGAACAAAGC	x	x
TCP1.75	GCAAGCCATACTCAAGGACAC	x	x
TCP1.76	CCAATCCTCTCCACGCTGCTC	x	x

TABLE 2
(cont.)

TCPI.77	GCGATGACCTCCGTGAGCCTG				X	X
TCPI.78	CCCAGGACAGGCTCACGGA				X	X
bilitCP1	CCTCTCAAGTGTGTGATTC				X	X
bilitCP2	CAGCTGACGACGTACACACTCATC				X	X
bilitCP4	CTGACGTCCAGACTCCGCTTCAT				X	X
bilitCP6	AGCTTGTTCCTCAATGTGCGGTAG				X	X
rpprim01	GACCTGAGCAGCTCGACGATACACACTCATC				X	X
L11	GTGTCGAGCTGCTCAGGTC				X	X
L12	AGCACGCTGAAACAGTCCCTT				X	X
L13	GACCTGAGCAGCTCGACGAC				X	X
L14	AAGGCACTGTTCAAGCGTCT				X	X
L15	CGGCGGAGTGTCTGGAGCMA				X	X
L16	GGAATGAAGCGGAGTCTGA				X	X
BamHI.L7	ATGGATCCGTCGTGAGCTGCTCAGGTCCT	BamHI site			X	X
SaIL.L8	ATCAGCTGAGCAGCGCTGAAACAGTGCCTTC	Pvu II site (not SaI I)			X	X
K303	GTCTCCGTGACATTAAGAAGAGAC				X	X
K304	GCCMAGTTCCTGCACCTGGCT				X	X
K305	GCCGTGTTCTTGAACCGTGGTCT				X	X
K306	XXGCCGTGTTCTTGAACCGTGGTCT	X - bhoI, ~ K305			X	X
K311	GTCAAGATGCTGAGATAGAAC				X	X
K312	TGCTTAGCTTGTGGGGGTGTCA				X	X
K313	TGCTTAGCTTGTGGGGGTGTCA				X	X
K320	GCTGGCTCCTGCTGGCGGACGT				X	X
K321	CAGCGGGAGCGCGCGGCAATC				X	X
K322	TGGGCCAACAGCGCGCGGAAA				X	X
slanti.1	CGGGCGCAAGCGCGTGGGG				X	X
slanti.2	CCGACAGCTCCCGCAGCTGCACCC				X	X
slanti.3	CGTACAGACTCATCAGCCAGTGCAGGAACTTGGC				X	X
slanti.4	CGGCGCCGCTCGTAGTTGAGCAGCGCTGAACAGTGCCTTC				X	X
slanti.5	GGCGAGTCTGAGACGTACAGAGGGCGGCGCTGGCTTCTTCG				X	X
UTR2	ATTTGACCCACAGGAGACCCCATCCAG				X	X
FW5	ATGACCGCCCTCCTCGTGAG				X	X
Nam1	GCCACCCCGCGATGCC				X	X
Nam2	AGCCCTGGCCCCGGCCA				X	X
Nam3	TCCACGTGCGCAGCAG				X	X
Nam4	AGCAGGACGACAGCGCTG				X	X
PE01	CGCGGTAGTGGTGGCGACGAGGAGGACGACGAC				X	X

TABLE 2
(cont.)

PE02	CCAGGGCTTCCACGCTGCGCAGGACGACCGACGGC	X	X	
LM101	CTAGTCTAGATCA / GCTAGCGTAATCTGGAAACATCTGA TGGGTA / GTCCAGGATGGTCTGAAGTC			X
LM103	TACCATGGGCTACCCATACGACGTTCCAGATTACGCTCA	inserts HA tag into a Nde I site at 5' end of hTRT		X
LM104	TATGAGCGTAATCTGGAAACGTCGTAATGGTAGCCCATGG	amends to LM103		X
LM105	GTGTACGTCGTCGAGCTCCTCAGGCTGCTGCTTTT	change = F559A (phe → ala)		X
LM106	ATGTCACGGAG GTGTACGTCGTCGAGCTCCTCAGGCTTTTGGCTTAATGTC	change = F560A (phe → ala)		X
LM107	ACGGAGACC CCTCAGGCTCTTTCTTTGCTGTCACGGAGACCAACGTTT	change = Y561A (tyr → ala)		X
LM108	CAMAGAACAG GGTCTTCTTTATATGTCGCGGAGACCAACGTTT	change = T561A (thr → ala)		X
LM109	CTTCTTTTATGTCACGGCGACACAGCTTTCAMAGAACACA	change = E564A deletion of FFVYTH (aa 559-564)		X
LM_FFYT	ATGAGTGTGTACGTCGTCGAGCTCCTCAGGCTTACGACG	complement to TCTPT61 to D12 motif, designed to	X	X
TCT061:	CAMAGAACAGGCTCTTTTTC GGCTGATGAGTGTGTACGTCGTCGA	possibly amend to mTRT	X	X
HUM01:	ACGTGCTCTCCGTGACATTAANAAGAA	possibly amend to mTRT	X	X
HUM02:	AGGCTTTTCTTTATATGTCACGGA	designed to	X	X
HUM03:	CACAGACCCCGTCGCTTGCTC	possibly amend to mTRT	X	X
HUM04:	CGGAGCTGTGACGTCAGCAGGGC	designed to	X	X
SLW FIN	cgcgaaacgclaaacaaATGCCGCGCGCTCCCGCTCTC	possibly amend to mTRT	X	X
SLW F1C	cgcgaaacgclaaacaaCAAAAGAGTGGCTTCTTCGGC	for GST fusion construct (782 to 1636) UC = hTRT seq, lc = BamHI site + 2 stop codons	X	X
SLW F2N	cgcgaaacgclaaacaaGCCACCTCTTTGGAAGGCTGCG	for GST fusion construct (782 to 1636) UC = hTRT seq, lc = EcoRI site + 3 stop codons	X	X
SLW F2C	cgcgaaacgclaaacaaAGACCTGAGCAGGCTCCGACGAC	for GST fusion construct (1625 to 2458) UC = hTRT seq, lc = EcoRI site + 3 stop codons	X	X
SLW F3N	SLW F2N / SLW F2C amplify a 872 nt piece of pGRN121 (1625 to 2458) cgcgaaacgclaaacaaATGAGTGTGTACGTCGTCGAC	for GST fusion construct (2426 to 3274)	X	X

TABLE 2
(cont.)

SLW F3C	<u>cgccgatlccglnclncln</u> GATCCCCCTGDCACTGGACG	UC = h1RT seq, le = BamHI site + 2 stop codons for GST fusion construct (2426 to 3274) UC = h1RT seq, le = EcoRI site + 3 stop codons	X	X
SLW F3N / SLW F3C	amplify a 887 nt piece of pGIRN121 (2426 to 3274)			
SLW F4N	<u>cgccgatlccglncln</u> gATCCCCGCAAGGCTCCATCCCTC	for GST fusion construct (3272 to 4177) UC = h1RT seq, le = BamHI site + 2 stop codons for GST fusion construct (3272 to 4177)	X	X
SLW F4C	<u>cgccgatlccglncln</u> clnGTCACAGGATTCCTCTGAAAGTTC	UC = h1RT seq, le = EcoRI site + 3 stop codons	X	X
SLW F4N / SLW F4C amplify a 944 nt piece of pGIRN121 (3272 to 4177)				
40-60	GGCATCCCGGGGGGCTGGCCGGG	phosphorothioate	X	X
260-280	GGACACCTGGCCGAAAGAGAGG	phosphorothioate	X	X
500-520	GCGTGCCAGCAGGTGAAACCAAG	phosphorothioate	X	X
770-790	CTCAGGGGCGACGCCGCCCT	phosphorothioate	X	X
885-905	AGGTGGCTTCTTCGGCGGGTC	phosphorothioate	X	X
1000-1020	GGACAAAGGCGTGTCCCAAGGA	phosphorothioate	X	X
1300-1320	GCTGGGGTGACCGGAGCTCGC	phosphorothioate	X	X
1520-1540	GATGAACCTTCTGGTGTTCCT	phosphorothioate	X	X
2110-2130	GTGGCCCAAGCCCTGTGGATA	phosphorothioate	X	X
2295-2315	GCCCATGGGGCGGCTTCTGGA	phosphorothioate	X	X
2450-2470	GAGGCCACTGCTGGCCATC	phosphorothioate	X	X
2670-2690	GGTGAAGTGAAGTGTACCA	phosphorothioate	X	X
3080-3110	GCTGCAGCACACATGCGTGAAACCTGTACGC	phosphorothioate	X	X
3140-3160	GACGGCGAGGAAAAATGTGGG	phosphorothioate	X	X
3690-3710	CCGAGCGCCAGGCTGTGGGA	phosphorothioate	X	X
55-75	CAGCGGGAGCGCGCGGCATC	phosphorothioate	X	X
151-171	CAGCACTCGCGGTAGTGGCT	phosphorothioate	X	X
TP1.1	TCAAGCCAAACCTGMACTGAG		X	
TP1.2	CCGAGTGAATCTTCTACGC		X	
TP1.3	GTCTCTGGCAGTTCTCATCCG		X	
TP1.4	TTAAGGCATCCTCCCAAGCACA		X	

For information about TRT proteins and peptides generally, TRT protein activities, telomerase catalytic activity (non-processive and processive), activity determination, other telomerase or TRT protein activities, the use of telomerase motifs as targets, chemical synthesis of hTRT and other TRT polypeptides, reference is hereby
5 made to parent UK Patent Application No 97 208890.4 published under Serial No 2317891.

IV) RECOMBINANT EXPRESSION OF hTRT AND OTHER TRT PROTEINS

10 The following section is generally allied to but not necessarily always directly describing the invention and concerns methods, reagents, vectors, and cells useful for expression of hTRT polypeptides and nucleic acids using *in vitro* (cell-free), *ex vivo* or *in vivo* (cell or organism-based) recombinant expression systems. Recombinant expression of the hTRT protein, or fragment thereof will usually involve inserting the
15 coding sequence into an appropriate expression vector (*i.e.*, a vector that contains the necessary elements for the transcription and translation of the inserted coding sequence required for the expression system employed). Suitable transcription units capable of expressing an hTRT polypeptide comprise a polynucleotide substantially identical in sequence to an hTRT gene coding sequence at least 25 nucleotides, preferably 50 to 100
20 nucleotides or more, of the hTRT cDNAs or gene operably linked to a promoter. Methods well known to those skilled in the art can be used to construct the expression vectors containing an hTRT sequence and appropriate transcriptional controls provided by the present invention (see, e.g., Sambrook et al., *supra*, Ausubel et al. *supra*, and this disclosure).

25 Recombinant hTRT polypeptides include fusion proteins that contain hTRT polypeptides or fragments of the hTRT protein. The fusion proteins are typically produced by recombinant means, although they may also be made by chemical synthesis. Fusion proteins can be useful in providing enhanced expression of the hTRT polypeptide constructs, or in producing hTRT polypeptides having other desirable
30 properties for example, comprising a label (such as an enzymatic reporter group), binding group, or antibody epitope. An exemplary fusion protein, comprising hTRT and

enhanced green fluorescent protein (EGFP) sequences is described in Example 15, *infra*. It will be apparent to one of skill that the uses and applications discussed in Example 15 and elsewhere herein are not limited to the particular fusion protein, but are illustrative of the uses of various fusion constructs.

5 Fusion protein systems can also be used to facilitate efficient production and isolation of hTRT proteins or peptides. For example, the non-hTRT sequence portion of the fusion protein comprises a short peptide that can be specifically bound to an immobilized molecule such that the fusion protein can be separated from unbound components (such as unrelated proteins in a cell lysate). One example is a peptide
10 sequence that is bound by a specific antibody. Another example is a peptide comprising polyhistidine tracts e.g. (His)₆ or histidine-tryptophan sequences that can be bound by a resin containing nickel or copper ions (i.e., metal-chelate affinity chromatography). Other examples include Protein A domains or fragments, which allow purification on immobilized immunoglobulin, and the domain utilized in the FLAGS extension/affinity
15 purification system (Immunex Corp, Seattle WA). The fusion protein may include a cleavage site so that the hTRT or other TRT polypeptide sequence can be easily separated from the non-hTRT peptide or protein sequence. In this case, cleavage may be chemical (e.g., cyanogen bromide, 2-(2-nitrophenylsulphenyl)-3-methyl-3'-bromoindolene, hydroxylamine, or low pH) or enzymatic (e.g., Factor Xa,
20 enterokinase). The choice of the fusion and cleavage systems may depend, in part, on the portion (i.e., sequence) of the hTRT polypeptide being expressed. Fusion proteins generally are described in Ausubel et al., *supra*, Ch. 16, Kroll et al., 1993, *DNA Cell Biol.* 12:441, and the Invitrogen 1997 Catalog (Invitrogen Inc, San Diego CA). Other exemplary fusion proteins with epitope tags or tags and cleavage sites are described in
25 Example 6, *infra*.

 It will be appreciated by those of skill that, although the expression systems discussed in this section are focused on expression of hTRT polypeptides, the same or similar cells, vectors and methods may be used to express hTRT polynucleotides, including sense and antisense polynucleotides without necessarily
30 desiring production of hTRT polypeptides. Typically, expression of a polypeptide requires a suitable initiation codon (e.g., methionine), open reading frame, and

translational regulatory signals (e.g., a ribosome binding site, a termination codon) which may be omitted when translation of a nucleic acid sequence to produce a protein is not desired.

Expression of hTERT polypeptides and polynucleotides may be carried out to accomplish any of several related benefits. One illustrative benefit is expression of hTERT polypeptides that are subsequently isolated from the cell in which they are expressed (for example for production of large amounts of hTERT for use as a vaccine or in screening applications to identify compounds that modulate telomerase activity). A second illustrative benefit is expression of hTERT in a cell to change the phenotype of the cell (as in gene therapy applications). Nonmammalian cells can be used for expression of hTERT for purification, while eukaryotic especially mammalian cells (e.g., human cells) can be used not only for isolation and purification of hTERT but also for expression of hTERT when a change in phenotype in a cell is desired (e.g., to effect a change in proliferative capacity as in gene therapy applications). By way of illustration, hTERT polypeptides having one or more telomerase activities (e.g. telomerase catalytic activity) can be expressed in a host cell to increase the proliferative capacity of a cell (e.g., immortalize a cell) and, conversely, hTERT antisense polynucleotides or inhibitory polypeptides typically can be expressed to reduce the proliferative capacity of a cell (e.g., of a telomerase positive malignant tumor cell). Numerous specific applications are possible as suggested, in the discussion of therapeutic applications, below.

Illustrative useful expression systems (cells, regulatory elements, vectors and expression) include a number of cell-free systems such as reticulocyte lysate and wheat germ systems using hTERT polynucleotides in accordance with general methods well known in the art (see, e.g., Ausubel et al. *supra* at Ch. 10). The present invention encompasses nucleic acids encoding hTERT polynucleotides, proteins, protein subsequences, or fusion proteins that can be expressed in prokaryotic or eukaryotic cells, eg bacteria, fungi, plant, insect, and animal, including human cell expression systems known in the art, including isolated cells, cell lines, cell cultures, tissues, and whole organisms. As will be understood by those of skill, hTERT polynucleotides introduced into a host cell or cell free expression system will be operably linked to appropriate expression control sequences for each host or cell free system.

Useful bacterial expression systems include *E. coli*, bacilli (such as *Bacillus subtilis*), other enterobacteriaceae (such as *Salmonella*, *Serratia*, and various *Pseudomonas species*) or other bacterial hosts (e.g., *Streptococcus cremoris*, *Streptococcus lactis*, *Streptococcus thermophilus*, *Leuconostoc citrovorum*,
5 *Leuconostoc mesenteroides*, *Lactobacillus acidophilus*, *Lactobacillus lactis*, *Bifidobacterium bifidum*, *Bifidobacteriu breve*, and *Bifidobacterium longum*). hTRT expression constructs useful in prokaryotes include recombinant bacteriophage, plasmid or cosmid DNA expression vectors, or the like, and typically include promoter sequences. Illustrative promoters include inducible promoters, such as the *lac* promoter,
10 the hybrid *lacZ* promoter of the Bluescript7 phagemid [Stratagene, La Jolla CA] or pSport1 [Gibco BRL]; phage lambda promoter systems; a tryptophan (*trp*) promoter system; and *p_{trp}-lac* hybrids and the like. Bacterial expression constructs may optionally include a ribosome binding site and transcription termination signal regulatory sequences. Illustrative examples of specific vectors useful for expression include, for
15 example, pTrcHis2, (Invitrogen, San Diego CA), pThioHis A, B & C, and numerous others known in the art or that may be developed (see, e.g. Ausubel). Useful vectors for bacteria include those that facilitate production of hTRT- fusion proteins. Useful vectors for high level expression of fusion proteins in bacterial cells include the multifunctional *E. coli* cloning and expression vectors such as Bluescript7 (Stratagene),
20 noted above, in which the sequence encoding hTRT protein, an hTRT fusion protein or an hTRT fragment may be ligated into the vector in-frame with sequences for the amino-terminal Met and the subsequent 7 residues of β -galactosidase so that a hybrid protein is produced (e.g., pIN vectors; Van Heeke and Schuster, 1989, *J. Biol. Chem.*, 264:5503). Vectors such as pGEX vectors (e.g., pGEX-2TK; Pharmacia Biotech) may
25 also be used to express foreign polypeptides, such as hTRT protein, as fusion proteins with glutathione S-transferase (GST). Such fusion proteins may be purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. Proteins made in such systems often include enterokinase, thrombin or factor Xa protease cleavage sites so that the cloned polypeptide of interest can be
30 released from the GST moiety at will, as may be useful in purification or other applications. Other examples are fusion proteins comprising hTRT and the *E. coli*

Maltose Binding Protein (MBP) or *E. Coli* thioredoxin. Illustrative examples of hTERT expression constructs useful in bacterial cells are provided in Example 6, *infra*.

hTERT polypeptides may be expressed in fungal systems, such as *Dictyostelium* and, preferably, yeast, such as *Saccharomyces cerevisiae*, *Pichia pastoris*, *Torulopsis holmii*, *Saccharomyces fragilis*, *Saccharomyces lactis*, *Hansenula polymorpha* and *Candida pseudotropicalis*. When hTERT is expressed in yeast, a number of suitable vectors are available, including plasmid and yeast artificial chromosomes (YACs) vectors. The vectors typically include expression control sequences, such as constitutive or inducible promoters (e.g., such as alpha factor, alcohol oxidase, PGH, and 3-phosphoglycerate kinase or other glycolytic enzymes), and an origin of replication, termination sequences and the like, as desired. Suitable vectors for use in *Pichia* include pPICZ, His6/pPICZB, pPICZalpha, pPIC3.5K, pPIC9K, pA0815, pGAP2A, B & C, pGAP2alpha A, B, and C (Invitrogen, San Diego, CA) and numerous others known in the art or to be developed. The vector His6/pPICZB (Invitrogen, San Diego, CA) is used to express a His₆-hTERT fusion protein in the yeast *Pichia pastoris*. An example of a vector useful in *Saccharomyces* is pYES2 (Invitrogen, San Diego, CA). Illustrative examples of hTERT expression constructs useful in yeast are provided in Example 6, *infra*.

hTERT polypeptides may also be expressed in plant cell systems transfected with plant or plant virus expression vectors (e.g., cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or transformed with bacterial expression vectors (e.g., Ti or pBR322 plasmid). In cases where plant virus expression vectors are used, the expression of an hTERT-encoding sequence may be driven by any of a number of promoters. For example, viral promoters such as the 35S and 19S promoters of CaMV (Brisson et al., 1984, *Nature* 310:511-514) may be used alone or in combination with the omega leader sequence from TMV (Takamatsu et al., 1987, *EMBO J.*, 6:307-311). Alternatively, plant promoters such as that from the small subunit gene of RUBISCO (Coruzzi et al., 1984, *EMBO J.*, 3:1671-1680; Broglie et al., 1984, *Science* 224:838-843) or heat shock promoters (Winter and Sinibaldi, 1991, *Results Probl. Cell Differ.*, 17:85), or storage protein gene promoters may be used. These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated

transfection (for reviews of such techniques, see Hobbs or Murry, 1992, in MCGRAW HILL YEARBOOK OF SCIENCE AND TECHNOLOGY McGraw Hill New York NY, pp. 191-196 [1992]; or Weissbach and Weissbach, 1988, METHODS FOR PLANT MOLECULAR BIOLOGY, Academic Press, New York NY, pp. 421-463).

5 Another expression system for expression of hTERT protein is an insect system. A preferred system uses a baculovirus polyhedrin promoter. In one such system, *Autographa californica* nuclear polyhedrosis virus (AcNPV) is used as a vector to express foreign genes in *Spodoptera frugiperda* cells or in *Trichoplusia* larvae. The sequence encoding the gene of interest may be cloned into a nonessential region of the virus, such as the polyhedrin gene, and placed under control of the polyhedrin promoter. 10 Successful insertion of the sequence, e.g., encoding the hTERT protein, will render the polyhedrin gene inactive and produce recombinant virus lacking coat protein. The recombinant viruses are then used to infect *S. frugiperda* cells or *Trichoplusia* larvae, in which the hTERT sequence is then expressed (see, for general methods, Smith et al., *J. Virol.*, 46:584 [1983]; Engelhard et al., *Proc. Natl. Acad. Sci.* 91:3224-7 [1994]). 15 Useful vectors for baculovirus expression include pBlueBacHis2 A, B & C, pBlueBac4.5, pMelBacB and numerous others known in the art or to be developed. Illustrative examples of hTERT expression constructs useful in insect cells are provided in Example 6, *infra*.

20 There are also expression systems in mammals and mammalian cells. As noted *supra*, hTERT polynucleotides may be expressed in mammalian cells (e.g., human cells) for production of significant quantities of hTERT polypeptides (e.g., for purification) or to change the phenotype of a target cell (e.g., for purposes of gene therapy, cell immortalization, or other). In the latter case, the hTERT polynucleotide 25 expressed may or may not encode a polypeptide with a telomerase catalytic activity. That is, expression may be of a sense or antisense polynucleotide, an inhibitory or stimulatory polypeptide, a polypeptide with zero, one or more telomerase activities, and other combinations and variants disclosed herein or apparent to one of skill upon review of this disclosure.

30 Suitable mammalian host tissue culture cells for expressing nucleic acids include any normal mortal or normal or abnormal immortal animal or human

cell, including: monkey kidney CV1 line transformed by SV40 (COS-7, ATCC CRL 1651); human embryonic kidney line (293; Graham et al., *J. Gen. Virol.* 36:59 (1977)); baby hamster kidney cells (BHK, ATCC CCL 10); CHO (ATCC CCL 61 and CRL 9618); mouse sertoli cells (TM4, Mather, *Biol. Reprod.* 23:243-251 (1980)); monkey kidney cells (CV1 ATCC CCL 70); African green monkey kidney cells (VERO-76, ATCC CRL 1587); human cervical carcinoma cells (HeLa, ATCC CCL 2); canine kidney cells (MDCK, ATCC CCL 34); buffalo rat liver cells (BRL 3A, ATCC CRL 1442); human lung cells (W138, ATCC CCL 75); human liver cells (Hep G2, HB 8065); mouse mammary tumor (MMT 060562, ATCC CCL51); TRI
5 cells (Mather, et al., *Annals N.Y. Acad. Sci.* 383:44-46 (1982); MDCK cells (ATCC CCL 34 and CRL 6253); HEK 293 cells (ATCC CRL 1573); and WI-38 cells (ATCC CCL 75; ATCC: American Type Culture Collection, Rockville, MD). The use of mammalian tissue cell culture to express polypeptides is discussed generally in Winnacker, *FROM GENES TO CLONES* (VCH Publishers, N.Y., N.Y., 1987).

15 For mammalian host cells, viral-based and nonviral expression systems are available. Nonviral vectors and systems include plasmids and episomal vectors, typically with an expression cassette for expressing a protein or RNA, and human artificial chromosomes (see, e.g., Harrington et al., 1997, *Nat Genet* 15:345). For example, nonviral vectors useful for expression of hTERT polynucleotides and
20 polypeptides in mammalian (e.g., human) cells include pcDNA3.1/His, pEBVHis A, B & C, (Invitrogen, San Diego CA), MPSV vectors, others described in the Invitrogen 1997 Catalog (Invitrogen Inc, San Diego CA), which is incorporated in its entirety herein, and numerous others known in the art for other proteins. Illustrative examples of hTERT expression constructs useful in mammalian cells are provided in Example 6,
25 *infra*.

Useful viral vectors include vectors based on retroviruses, adenoviruses, adenoassociated viruses, herpes viruses, vectors based on SV40, papilloma virus, HBP Epstein Barr virus, vaccinia virus vectors and Semliki Forest virus (SFV). SFV and vaccinia vectors are discussed generally in Ausubel et al., *supra*, Ch 16. These vectors
30 are often made up of two components, a modified viral genome and a coat structure surrounding it (see generally Smith, 1995, *Annu. Rev. Microbiol.* 49: 807), although

sometimes viral vectors are introduced in naked form or coated with proteins other than viral proteins. However, the viral nucleic acid in a vector may be changed in many ways, for example, when designed for gene therapy. The goals of these changes are to disable growth of the virus in target cells while maintaining its ability to grow in vector form in available packaging or helper cells, to provide space within the viral genome for insertion of exogenous DNA sequences, and to incorporate new sequences that encode and enable appropriate expression of the gene of interest. Thus, vector nucleic acids generally comprise two components: essential cis-acting viral sequences for replication and packaging in a helper line and the transcription unit for the exogenous gene. Other viral functions are expressed in trans in a specific packaging or helper cell line.

Adenoviral vectors (e.g., for use in human gene therapy) are described in, e.g., Rosenfeld et al., 1992, *Cell* 68: 143; PCT publications WO 94/12650; 94/12649; and 94/12629. In cases where an adenovirus is used as an expression vector, a sequence encoding hTRT may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a nonessential E1 or E3 region of the viral genome will result in a viable virus capable of expressing in infected host cells (Logan and Shenk, 1984, *Proc. Natl. Acad. Sci.*, 81:3655).

Replication-defective retroviral vectors harboring a therapeutic polynucleotide sequence as part of the retroviral genome are described in, e.g., Miller et al., 1990, *Mol. Cell Biol.* 10: 4239; Kolberg, 1992, *J. NIH Res.* 4: 43; and Cornetta et al., 1991, *Hum. Gene Ther.* 2: 215.

In mammalian cell systems, promoters from mammalian genes or from mammalian viruses are often appropriate. Suitable promoters may be constitutive, cell type-specific, stage-specific, and/or modulatable or regulatable (e.g., by hormones such as glucocorticoids). Useful promoters normally include the metallothionein promoter, the constitutive adenovirus major late promoter, the dexamethasone-inducible MMTV promoter, the SV40 promoter, the MRP polIII promoter, the constitutive MPSV promoter, the tetracycline-inducible CMV promoter (such as the human immediate-early CMV promoter), the constitutive CMV promoter, and promoter-enhancer combinations known in the art.

Other regulatory elements may also be required or desired for efficient

expression of an hTERT polynucleotide and/or translation of a sequence encoding hTERT proteins. For translation, these elements typically include an ATG initiation codon and adjacent ribosome binding site or other sequences. For sequences encoding the hTERT protein, provided its initiation codon and upstream promoter sequences of the invention
5 are inserted into an expression vector, no additional translational or other control signals may be needed.

Expression of hTERT gene products can also be effected (increased) by activation of an hTERT promoter or enhancer in a cell such as a human cell, e.g., a telomerase-negative cell line. Activation can be carried out in a variety of ways,
10 including administration of an exogenous promoter activating agent, or inhibition of a cellular component that suppresses expression of the hTERT gene. It will be appreciated that, conversely, inhibition of promoter function, as described *infra*, will reduce hTERT gene expression.

Inducible and repressible expression of hTERT polypeptides is possible
15 using such systems as the Ecdysone-Inducible Expression System (Invitrogen), and the Tet-On and Tet-off tetracycline regulated systems from Clontech. The ecdysone-inducible expression system uses the steroid hormone ecdysone analog, muristerone A, to activate expression of a recombinant protein via a heterodimeric nuclear receptor (No et al., 1996, *Proc. Natl. Acad. Sci. USA* 93:3346). hTERT can be
20 cloned in the pIND vector (Clontech), which contains five modified ecdysone response elements (E/GREs) upstream of a minimal heat shock promoter and the multiple cloning site. The construct is then transfected in cell lines stably expressing the ecdysone receptor. After transfection, cells are treated with muristerone A to induce intracellular expression from pIND. Alternatively, hTERT polypeptide may be expressed using the
25 Tet-on and Tet-off expression systems (Clontech) to provide regulated, high-level gene expression (Gossen et al., 1992, *Proc. Natl. Acad. Sci. USA* 89:5547; Gossen et al., 1995, *Science* 268:1766).

The hTERT vectors of the invention may be introduced into a cell, tissue, organ, patient or animal by a variety of methods. The nucleic acid expression vectors
30 (typically dsDNA) of the invention can be transferred into the chosen host cell by well-known methods such as calcium chloride transformation (for bacterial systems),

electroporation, calcium phosphate treatment, liposome-mediated transformation, injection and microinjection, ballistic methods, virosomes, immunoliposomes, polycation:nucleic acid conjugates, naked DNA, artificial virions, fusion to the herpes virus structural protein VP22 (Elliot and O'Hare, *Cell* 88:223), agent-enhanced uptake of DNA, and *ex vivo* transduction. Useful liposome-mediated DNA transfer methods are described in US Patent Nos. 5,049,386, US 4,946,787, and US 4,897,355; PCT publications WO 91/17424, WO 91/16024; Wang and Huang, 1987, *Biochem. Biophys. Res. Commun.* 147: 980; Wang and Huang, 1989, *Biochemistry* 28: 9508; Litzinger and Huang, 1992, *Biochem. Biophys. Acta* 1113:201; Gao and Huang, 1991, *Biochem. Biophys. Res. Commun.* 179: 280. Immunoliposomes have been described as carriers of exogenous polynucleotides (Wang and Huang, 1987, *Proc. Natl. Acad. Sci. U.S.A.* 84:7851; Trubetskoy et al., 1992, *Biochem. Biophys. Acta* 1131:311) and may have improved cell type specificity as compared to liposomes by virtue of the inclusion of specific antibodies which presumably bind to surface antigens on specific cell types. Behr et al., 1989, *Proc. Natl. Acad. Sci. U.S.A.* 86:6982 report using lipopolyamine as a reagent to mediate transfection itself, without the necessity of any additional phospholipid to form liposomes. Suitable delivery methods will be selected by practitioners in view of acceptable practices and regulatory requirements (e.g., for gene therapy or production of cell lines for expression of recombinant proteins). It will be appreciated that the delivery methods listed above may be used for transfer of nucleic acids into cells for purposes of gene therapy, transfer into tissue culture cells, and the like.

For long-term, high-yield production of recombinant proteins, stable expression will often be desired. For example, cell lines which stably express hTERT can be prepared using expression vectors of the invention which contain viral origins of replication or endogenous expression elements and a selectable marker gene. Following the introduction of the vector, cells may be allowed to grow for 1-2 days in an enriched media before they are switched to selective media. The purpose of the selectable marker is to confer resistance to selection, and its presence allows growth of cells which successfully express the introduced sequences in selective media. Resistant, stably transfected cells can be proliferated using tissue culture techniques appropriate to the

cell type. An amplification step, e.g., by administration of methyldtrexate to cells transfected with a DHFR gene according to methods well known in the art, can be included.

In addition, a host cell strain may be chosen for its ability to modulate the expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, phosphorylation, lipidation and acylation. Post-translational processing may also be important for correct insertion, folding and/or function. Different host cells have cellular machinery and characteristic mechanisms specific for each cell for such post-translational activities and so a particular cell may be chosen to ensure the correct modification and processing of the introduced, foreign protein.

The present invention also provides transgenic animals (i.e., mammals transgenic for a human or other TRT gene sequence) expressing an hTRT or other TRT polynucleotide or polypeptide under the control of an hTRT promoter sequence of the invention. In one embodiment, hTRT is secreted into the milk of a transgenic mammal such as a transgenic bovine, goat, or rabbit. Methods for production of such animals are found, e.g., in Heyneker et al., PCT WO 91/08216.

The hTRT proteins and complexes arising out of the invention, including those made using expression systems disclosed herein *supra*, may be purified using a variety of general methods known in the art in accordance with the specific methods described in parent UK Patent Application No 97 20890.4 published under Serial No 2317891. One of skill in the art will recognize that after chemical synthesis, biological expression, or purification, the hTRT protein may possess a conformation different than a native conformation of naturally occurring telomerase. In some instances, it may be helpful or even necessary to denature (e.g., including reduction of disulfide or other linkages) the polypeptide and then to cause the polypeptide to re-fold into the preferred conformation. Productive refolding may also require the presence of hTR (or hTR fragments). Methods of reducing and denaturing proteins and inducing re-folding are well known to those of skill in the art (see, e.g., Debinski et al., 1993, *J. Biol. Chem.*, 268:14065; Kreitman and Pastan, 1993, *Bioconjug. Chem.*, 4:581; and Buchner et al., 1992, *Anal. Biochem.*, 205:263; and McCaman et al., 1985, *J. Biotech.* 2:177). See

also PCT Publication WO 96/40868, *supra*.

**V) COMPLEXES OF HUMAN TRT AND HUMAN TELOMERASE
RNA, TELOMERASE-ASSOCIATED PROTEINS, AND OTHER
5 BIOMOLECULES PRODUCED BY COEXPRESSION AND OTHER MEANS**

Although not directly of the invention, the following description is of relevance to embodiments of the invention, for example if an hTRT promoter of the invention is employed in the recombinant expression of an hTRT protein.

hTRT polypeptides can associate *in vivo* and *in vitro* with other
10 biomolecules, including RNAs (e.g., hTR), proteins (e.g., telomerase-associated proteins), DNA (e.g., telomeric DNA, $[T_2AG_3]_N$), and nucleotides, such as (deoxy)ribonucleotide triphosphates. These associations can be exploited to assay hTRT presence or function, to identify or purify hTRT or telomerase-associated molecules, and to analyze hTRT or telomerase structure or function.

15 hTRT complexes with (e.g., associates with or binds to) a nucleic acid, usually an RNA, for example to produce a telomerase holoenzyme. The bound RNA is capable of acting as a template for telomerase-mediated DNA synthesis. Examples of RNAs that may be complexed with the hTRT polypeptide include a naturally occurring host cell telomerase RNA, a human telomerase RNA (e.g., hTR; U.S. Patent No.
20 5,583,016), an hTR subsequence or domain, a synthetic RNA, or other RNAs. The RNA-hTRT protein complex (an RNP) typically exhibits one or more telomerase activities, such as telomerase catalytic activities. These hTRT-hTR RNPs (or other hTRT-RNA complexes) can be produced by a variety of methods, as described *infra* for illustrative purposes, including *in vitro* reconstitution, by co-expression of hTRT and
25 hTR (or other RNA) *in vitro* (i.e., in a cell free system), *in vivo* reconstitution, or *ex vivo* reconstitution.

Thus, an hTRT-hTR complex (or other hTRT-RNA complex) may be formed *in vitro* by mixing separately purified components ("*in vitro* reconstitution," see, e.g., U.S. Patent No. 5,583,016 for a description of reconstitution; also see Autexier et
30 al., *EMBO J.* 15:5928).

Telomerase RNPs may be produced by coexpression of the hTRT

polypeptide and an RNA (e.g., hTR) *in vitro* in a cell-free transcription-translation system (e.g. wheat germ or rabbit reticulocyte lysate). As shown in Example 7, *in vitro* co-expression of a recombinant hTRT polypeptide and hTR results in production of telomerase catalytic activity (as measured by a TRAP assay).

5 Telomerase RNPs produced by expression of the hTRT polypeptide in a cell, e.g., a mammalian cell, in which hTR is naturally expressed or in which hTR (or another RNA capable of forming a complex with the hTRT protein) may be introduced or expressed by recombinant means. Thus, hTRT may be expressed in a telomerase negative human cell in which hTR is present (e.g., BJ or IMP90 cells), allowing the two
10 molecules to assemble into an RNP. Alternatively, hTRT may be expressed in a human or non-human cell in which hTR is recombinantly expressed. Methods for expression of hTR in a cell are found in U.S. Patent 5,583,016. Further, a clone containing a cDNA encoding the RNA component of telomerase has been placed on deposit as pGRN33 (ATCC 75926). Genomic sequences encoding the RNA component of human
15 telomerase are also on deposit in the ~15 kb *Sau*III A1 to *Hind*III insert of lambda clone 28-1 (ATCC 75925). For expression in eukaryotic cells the hTRT sequence will typically be operably linked to a transcription initiation sequence (RNA polymerase binding site) and transcription terminator sequences (see, e.g., PCT Publication WO 96/01835; Feng et al., 1995, *Science* 269:1236).

20 Recombinantly produced or substantially purified hTRT polypeptides may be coexpressed and/or associated with so-called "telomerase-associated proteins." Thus, the present invention provides hTRT coexpressed with, or complexed with, other proteins (e.g., telomerase-associated proteins). Telomerase-associated proteins are those proteins that copurify with human telomerase and/or that may play a role in
25 modulating telomerase function or activity, for example by participating in the association of telomerase with telomeric DNA. Examples of telomerase-associated proteins include (but are not limited to) the following proteins and/or their human homologs: nucleolin (see, Srivastava et al., 1989, *FEBS Letts.* 250:99); EF2H (elongation factor 2 homolog; see Nomura et al. 1994, *DNA Res. (Japan)* 1:27,
30 GENBANK accession #D21163); TP1/TLP1 (Harrington et al., 1997, *Science* 275:973; Nakayama, 1997, *Cell* 88:875); the human homologue of the *Tetrahymena* p95 or p95

itself (Collins et al., 1995, *Cell* 81:677); TPC2 (a telomere length regulatory protein; ATCC accession number 97708; TPC3 (also a telomere length regulatory protein; ATCC accession number 97707; DNA-binding protein B (dbpB; Horwitz et al., 1994, *J. Biol. Chem.* 269:14130; and Telomere Repeat Binding Factors (TRF 1 & 2; Chang et al., 1995, *Science* 270:1663; Chong et al., 1997, *Hum Mol Genet* 6:69); EST1, 3 and 4 (Lendvay et al., 1996, *Genetics* 144:1399, Nugent et al., 1996, *Science* 274:249, Lundblad et al., 1989, *Cell* 57:633); and End-capping factor (Cardenas et al., 1993, *Genes Dev.* 7:883).

Telomerase associated proteins can be identified on the basis of co-purification with, or binding to, hTERT protein or the hTERT-hTR RNP. Alternatively, they can be identified on the basis of binding to an hTERT fusion protein, e.g., a GST-hTERT fusion protein or the like, as determined by affinity purification (see, Ausubel et al. Ch 20). A particularly useful technique for assessing protein-protein interactions, which is applicable to identifying hTERT-associated proteins, is the two hybrid screen method of Chien et al. (*Proc. Natl. Acad. Sci. USA* 88:9578 [1991]; see also Ausubel et al., *supra*, at Ch. 20). This screen identifies protein-protein interactions *in vivo* through reconstitution of a transcriptional activator, the yeast Gal4 transcription protein (see, Fields and Song, 1989, *Nature* 340:245. The method is based on the properties of the yeast Gal4 protein, which consists of separable domains responsible for DNA-binding and transcriptional activation. Polynucleotides, usually expression vectors, encoding two hybrid proteins are constructed. One polynucleotide comprises the yeast Gal4 DNA-binding domain fused to a polypeptide sequence of a protein to be tested for an hTERT interaction (e.g., nucleolin or EF2H). Alternatively the yeast Gal4 DNA-binding domain is fused to cDNAs from a human cell, thus creating a library of human proteins fused to the Gal4 DNA binding domain for screening for telomerase associated proteins. The other polynucleotide comprises the Gal4 activation domain fused to an hTERT polypeptide sequence. The constructs are introduced into a yeast host cell. Upon expression, intermolecular binding between hTERT and the test protein can reconstitute the Gal4 DNA-binding domain with the Gal4 activation domain. This leads to the transcriptional activation of a reporter gene (e.g., lacZ, HIS3) operably linked to a Gal4 binding site. By selecting for, or by assaying the reporter, gene colonies of cells that

contain an hTRT interacting protein or telomerase associated protein can be identified. Those of skill will appreciate that there are numerous variations of the 2-hybrid screen, e.g., the LexA system (Bartel et al, 1993, *in Cellular Interactions in Development: A Practical Approach* Ed. Hartley, D.A. (Oxford Univ. Press) pp. 153-79).

- 5 Another useful method for identifying telomerase-associated proteins is a three-hybrid system (see, e.g., Zhang et al., 1996, *Anal. Biochem.* 242:68; Licitra et al., 1996, *Proc. Natl. Acad. Sci. USA* 93:12817). The telomerase RNA component can be utilized in this system with the TRT or hTRT protein and a test protein. Another useful method for identifying interacting proteins, particularly (i.e., proteins that
- 10 heterodimerize or form higher order heteromultimers), is the *E. coli*/BCCP interactive screening system (see, Germino et al. (1993) *Proc. Natl. Acad. Sci. U.S.A.* 90:933; Guarente (1993) *Proc. Natl. Acad. Sci. (U.S.A.)* 90:1639).

- Complexes of telomere binding proteins (which may or may not be telomerase associated proteins) and hTRT (which may or may not be complexed with
- 15 hTR, other RNAs, or one or more telomerase associated proteins) can arise. Examples of telomere binding proteins include TRF1 and TRF2 (*supra*); rnpA1, rnpA2, RAP1 (Buchman et al., 1988, *Mol. Cell. Biol.* 8:210, Buchman et al., 1988, *Mol. Cell. Biol.* 8:5086), SIR3 and SIR4 (Aparicio et al, 1991, *Cell* 66:1279), TEL1 (Greenwell et al., 1995, *Cell* 82:823; Morrow et al., 1995, *Cell* 82:831); ATM (Savitsky et al., 1995,
- 20 Science 268:1749), end-capping factor (Cardenas et al., 1993, *Genes Dev.* 7:883), and corresponding human homologs. The aforementioned complexes may be produced generally as described *supra* for complexes of hTRT and hTR or telomerase associated proteins, e.g., by mixing or co-expression *in vitro* or *in vivo*.

- Allied to the invention but not part of it are antibodies and other binding
- 25 agents reactive with hTRT. For more details about these and in particular monoclonal antibodies, human antibodies, humanised or chimeric antibodies, phage display, hybrid antibodies, anti-idotypic antibodies and antibodies generally, reference is made to parent UK Patent Application No 97 208890.4 published under Serial No 2317891.

- As regards purification of human telomerase reference is again made to
- 30 parent UK Patent Application No 97 208890.4 published under Serial No 2317891.

VI) TREATMENT OF TELOMERASE-RELATED DISEASE

A) INTRODUCTION

The present invention provides hTERT polynucleotides useful for the treatment of human diseases and disease conditions. Recombinant and synthetic hTERT gene products (protein and mRNA) can be used to create or elevate telomerase activity in a cell, as well as to inhibit telomerase activity in cells in which it is not desired. Thus, inhibiting, activating or otherwise altering a telomerase activity (e.g., telomerase catalytic activity, fidelity, processivity, telomere binding, *etc.*) in a cell can be used to change the proliferative capacity of the cell. For example, reduction of telomerase activity in an immortal cell, such as a malignant tumor cell, can render the cell mortal. Conversely, increasing the telomerase activity in a mortal cell (e.g., most human somatic cells) can increase the proliferative capacity of the cell. For example, expression of hTERT protein in dermal fibroblasts, thereby increasing telomere length, will result in increased fibroblast proliferative capacity; such expression can slow or reverse the age-dependent slowing of wound closure (see, e.g., West, 1994, *Arch. Derm.* 130:87).

Thus, in one aspect, the present invention provides reagents and methods useful for treating diseases and conditions characterized by the presence, absence, or amount of human telomerase activity in a cell and that are susceptible to treatment using the compositions and methods disclosed herein. These diseases include, as described more fully below, cancers, other diseases of cell proliferation (particularly diseases of aging), immunological disorders, infertility (or fertility), and others.

B) TREATMENT OF CANCER

The present invention provides methods and compositions for reducing telomerase activity in tumor cells and for treating cancer. Compositions include antisense oligonucleotides, gene therapy vectors encoding antisense oligonucleotides or activity altering proteins. Cancer cells (e.g., malignant tumor cells) that express telomerase activity (telomerase-positive cells) can be mortalized by decreasing or inhibiting the endogenous telomerase activity. Moreover, because telomerase levels correlate with disease characteristics such as metastatic potential (e.g., U.S. Patent No. 5,639,613; 5,648,215; 5,489,508; Pandita et al., 1996, *Proc. Am. Ass. Cancer Res.*

37:559), any reduction in telomerase activity could reduce the aggressive nature of a cancer to a more manageable disease state (increasing the efficacy of traditional interventions).

The invention provides compositions and methods useful for treatment of

5 cancers of any of a wide variety of types, including solid tumors and leukemias. Types of cancer that may be treated include (but are not limited to): adenocarcinoma of the breast, prostate, and colon; all forms of bronchogenic carcinoma of the lung; myeloid; melanoma; hepatoma; neuroblastoma; papilloma; apudoma; choristoma; branchioma; malignant carcinoid syndrome; carcinoid heart disease; carcinoma (e.g., Walker, basal

10 cell, basosquamous, Brown-Pearce, ductal, Ehrlich tumor, in situ, Krebs 2, merkel cell, mucinous, non-small cell lung, oat cell, papillary, scirrhous, bronchiolar, bronchogenic, squamous cell, and transitional cell), histiocytic disorders; leukemia (e.g., B-cell, mixed-cell, null-cell, T-cell, T-cell chronic, HTLV-II-associated, lymphocytic acute, lymphocytic chronic, mast-cell, and myeloid); histiocytosis malignant; Hodgkin's disease;

15 immunoproliferative small; non-Hodgkin's lymphoma; plasmacytoma; reticuloendotheliosis; melanoma; chondroblastoma; chondroma; chondrosarcoma; fibroma; fibrosarcoma; giant cell tumors; histiocytoma; lipoma; liposarcoma; mesothelioma; myxoma; myxosarcoma; osteoma; osteosarcoma; Ewing's sarcoma; synovioma; adenofibroma; adenolymphoma; carcinosarcoma; chordoma;

20 craniopharyngioma; dysgerminoma; hamartoma; mesenchymoma; mesonephroma; myosarcoma; ameloblastoma; cementoma; odontoma; teratoma; thymoma; trophoblastic tumor; adenocarcinoma; adenoma; cholangioma; cholesteatoma; cylindroma; cystadenocarcinoma; cystadenoma; granulosa cell tumor; gynandroblastoma; hepatoma; hidradenoma; islet cell tumor; leydig cell tumor; papilloma; sertoli cell tumor; theca cell

25 tumor; leiomyoma; leiomyosarcoma; myoblastoma; myoma; myosarcoma; rhabdomyoma; rhabdomyosarcoma; ependymoma; ganglioneuroma; glioma; medulloblastoma; meningioma; neurilemmoma; neuroblastoma; neuroepithelioma; neurofibroma; neuroma; paraganglioma; paraganglioma nonchromaffin; angiokeratoma; angiolymphoid hyperplasia with eosinophilia; angioma sclerosing; angiomatosis;

30 glomangioma; hemangioendothelioma; hemangioma; hemangiopericytoma; hemangiosarcoma; lymphangioma; lymphangiomyoma; lymphangiosarcoma; pinealoma;

carcinosarcoma; chondrosarcoma; cystosarcoma phyllodes; fibrosarcoma; hemangiosarcoma; leiomyosarcoma; leukosarcoma; liposarcoma; lymphangiosarcoma; myosarcoma; myxosarcoma; ovarian carcinoma; rhabdomyosarcoma; sarcoma (e.g., Ewing's, experimental, Kaposi's, and mast-cell); neoplasms (e.g., bone, breast, digestive system, colorectal, liver, pancreatic, pituitary, testicular, orbital, head and neck, central nervous system, acoustic, pelvic, respiratory tract, and urogenital); neurofibromatosis, and cervical dysplasia). The invention provides compositions and methods useful for treatment of other conditions in which cells have become immortalized or hyperproliferative, e.g., by dysregulation (e.g., abnormally high expression) of hTERT, telomerase enzyme, or telomerase activity.

The present invention further provides compositions and methods for prevention of cancers, gene therapy vectors that prevent telomerase activation, and gene therapy vectors that result in specific death of telomerase-positive cells. In a related aspect, the gene replacement therapy methods described below may be used for "treating" a genetic predilection for cancers.

C) TREATMENT OF OTHER CONDITIONS

The present invention also provides compositions and methods useful for treatment of diseases and disease conditions (in addition to cancers) characterized by under- or over-expression of telomerase or hTERT gene products. Examples include: diseases of cell proliferation, diseases resulting from cell senescence (particularly diseases of aging), immunological disorders, infertility, diseases of immune dysfunction, and others.

Certain diseases of aging are characterized by cell senescence-associated changes due to reduced telomere length (compared to younger cells), resulting from the absence (or much lower levels) of telomerase activity in the cell. Decreased telomere length and decreased replicative capacity contribute to diseases such as those described below. Telomerase activity and telomere length can be increased by, for example, increasing levels of hTERT gene products (protein and mRNA) in the cell. A partial listing of conditions associated with cellular senescence in which hTERT expression can be therapeutic includes Alzheimer's disease, Parkinson's disease, Huntington's disease,

products in the cell can be increased. Telomerase activity in a cell can also be increased by interfering with the interaction of endogenous telomerase inhibitors and the telomerase RNP, or endogenous hTERT transcription repressors and the hTERT gene; by increasing expression or activity of hTERT transcription activators; and other means
5 apparent to those of skill upon review of this disclosure.

E) INTERVENTION AGENTS

TRT PROTEINS & PEPTIDES

Telomerase modulatory polypeptides (i.e., proteins, polypeptides, and
10 peptides) that increase or reduce telomerase activity which can be expressed in a cell by introducing a nucleic acid (e.g., a DNA expression vector or mRNA) encoding the desired protein or peptide into the cell. Expression may be either constitutive or inducible depending on the vector and choice of promoter (*see* discussion below). Methods for introduction and expression of nucleic acids into a cell are well known in
15 the art (also, see elsewhere in this specification, e.g., sections on oligonucleotides, gene therapy methods).

Various telomerase modulatory and inhibitory peptides are described in parent UK Patent Application No 97 208890.4 published under Serial No 2317891.

In one embodiment of the invention, expression of the endogenous hTERT
20 gene is repressed by introduction into the cell of a large amount of hTERT polypeptide (e.g., typically at least about 2-fold more than the endogenous level, more often at least about 10- to about 100-fold) which acts via a feedback loop to inhibit transcription of the hTERT gene, processing of the hTERT pre-mRNA, translation of the hTERT mRNA, or assembly and transport of the telomerase RNP.

25

2) OLIGONUCLEOTIDES

a) ANTISENSE CONSTRUCTS

The invention provides methods and antisense oligonucleotide or polynucleotide reagents which can be used to reduce expression of hTERT gene products
30 *in vitro* or *in vivo*. Administration of the antisense reagents of the invention to a target cell results in reduced telomerase activity, and is particularly useful for treatment of

invention provide means for effecting immunosuppression. Conversely, the methods and reagents of the invention are useful for increasing telomerase activity and proliferative potential in cells, such as stem cells, that express a low level of telomerase or no telomerase prior to therapeutic intervention.

5

D) MODES OF INTERVENTION

As is clear from the foregoing discussion, modulation of the level of telomerase or telomerase activity of a cell can have a profound effect on the proliferative potential of the cell, and so has great utility in treatment of disease. As is also clear, this modulation may be either a decrease in telomerase activity or an increase in activity. The telomerase modulatory molecules of the invention can act through a number of mechanisms; some of these are described in this and the following subsections to aid the practitioner in selecting therapeutic agents. However, applicants do not intend to be limited to any particular mechanism of action for the novel therapeutic compounds, compositions and methods described herein.

Telomerase activity may be decreased through any of several mechanisms or combinations of mechanisms. One mechanism is the reduction of hTERT gene expression to reduce telomerase activity. This reduction can be at the level of transcription of the hTERT gene into mRNA.

Other mechanisms are referred to in parent UK Patent Application No 97 208890.4 published under Serial No 2317891.

In a one mechanism, an hTERT promoter sequence is operably linked to a gene encoding a toxin and introduced into a cell; if or when hTERT transcriptional activators are expressed or activated in the cell, the toxin will be expressed, resulting in specific cell killing.

Telomerase activity may be increased through any of several mechanisms, or a combination of mechanisms. These include increasing the amount of hTERT in a cell. Usually this is carried out by introducing an hTERT polypeptide-encoding polynucleotide into the cell (e.g., a recombinantly produced polypeptide comprising an hTERT DNA sequence operably linked to a promoter of the invention. In other mechanisms, expression from the endogenous hTERT gene or the stability of hTERT gene

relatively accessible sequences of the hTERT mRNA (e.g., relatively devoid of secondary structure). This can be determined by analyzing predicted RNA secondary structures using, for example, the MFOLD program (Genetics Computer Group, Madison WI) and testing *in vitro* or *in vivo* as is known in the art. Examples of oligonucleotides that may be tested in cells for antisense suppression of hTERT function are those capable of hybridizing to (i.e., substantially complementary to) the following positions from SEQUENCE ID NO:1: 40-60; 260-280; 500-520; 770-790; 885-905; 1000-1020 ; 1300-1320; 1520-1540; 2110-2130; 2295-2315; 2450-2470; 2670-2690; 3080-3110; 3140-3160; and 3690-3710. Another useful method for identifying effective antisense compositions uses combinatorial arrays of oligonucleotides (see, e.g., Milner et al., 1997, *Nature Biotechnology* 15:537). The invention also provides an antisense polynucleotide that has sequences in addition to the antisense sequence (i.e., in addition to anti-hTERT-sense sequence). In this case, the antisense sequence is contained within a polynucleotide of longer sequence. In another embodiment, the sequence of the polynucleotide consists essentially of, or is, the antisense sequence.

The antisense nucleic acids (DNA, RNA, modified, analogues, and the like) can be made using any suitable method for producing a nucleic acid, such as the chemical synthesis and recombinant methods disclosed herein. In one embodiment, for example, antisense RNA molecules of the invention may be prepared by *de novo* chemical synthesis or by cloning. For example, an antisense RNA that hybridizes to hTERT mRNA can be made by inserting (ligating) an hTERT DNA sequence (e.g., SEQUENCE ID No; 1, or fragment thereof) in reverse orientation operably linked to a promoter in a vector (e.g., plasmid). Provided that the promoter and, preferably termination and polyadenylation signals, are properly positioned, the strand of the inserted sequence corresponding to the noncoding strand will be transcribed and act as an antisense oligonucleotide of the invention.

The antisense oligonucleotides of the invention can be used to inhibit telomerase activity in cell-free extracts, cells, and animals, including mammals and humans. For example, the phosphorothioate antisense oligonucleotides:

- 30 A) 5'-GGCATCGCGGGGGTGGCCGGG
 B) 5'-CAGCGGGGAGCGCGCGGCATC
 C) 5'-CAGCACCTCGCGGTAGTGGCT'

diseases characterized by high telomerase activity (e.g., cancers). Without intending to be limited to any particular mechanism, it is believed that antisense oligonucleotides bind to, and interfere with the translation of, the sense hTERT mRNA. Alternatively, the antisense molecule may render the hTERT mRNA susceptible to nuclease digestion, interfere with transcription, interfere with processing, localization or otherwise with RNA precursors ("pre-mRNA"), repress transcription of mRNA from the hTERT gene, or act through some other mechanism. However, the particular mechanism by which the antisense molecule reduces hTERT expression is not critical.

The antisense polynucleotides of the invention comprise an antisense sequence of at least 7 to 10 to typically 20 or more nucleotides that specifically hybridize to a sequence from mRNA encoding hTERT or mRNA transcribed from the hTERT gene. More often, the antisense polynucleotide of the invention is from about 10 to about 50 nucleotides in length or from about 14 to about 35 nucleotides in length. In other embodiments, antisense polynucleotides are polynucleotides of less than about 100 nucleotides or less than about 200 nucleotides. In general, the antisense polynucleotide should be long enough to form a stable duplex but short enough, depending on the mode of delivery, to administer *in vivo*, if desired. The minimum length of a polynucleotide required for specific hybridization to a target sequence depends on several factors, such as G/C content, positioning of mismatched bases (if any), degree of uniqueness of the sequence as compared to the population of target polynucleotides, and chemical nature of the polynucleotide (e.g., methylphosphonate backbone, peptide nucleic acid, phosphorothioate), among other factors.

Generally, to assure specific hybridization, the antisense sequence is substantially complementary to the target hTERT mRNA sequence. In certain embodiments, the antisense sequence is exactly complementary to the target sequence. The antisense polynucleotides may also include, however, nucleotide substitutions, additions, deletions, transitions, transpositions, or modifications, or other nucleic acid sequences or non-nucleic acid moieties so long as specific binding to the relevant target sequence corresponding to hTERT RNA or its gene is retained as a functional property of the polynucleotide.

In one embodiment, the antisense sequence is complementary to

b) TRIPLEX OLIGO- AND POLYNUCLEOTIDES

The present invention provides oligo- and polynucleotides (e.g., DNA, RNA, PNA or the like) that bind to double-stranded or duplex hTERT nucleic acids; namely the hTERT gene to form a triple helix-containing, or "triplex" nucleic acid. Triple helix formation results in inhibition of hTERT expression by, for example, preventing transcription of the hTERT gene, thus reducing or eliminating telomerase activity in a cell. Without intending to be bound by any particular mechanism, it is believed that triple helix pairing compromises the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules to occur.

Triplex oligo- and polynucleotides of the invention are constructed using the base-pairing rules of triple helix formation (see, e.g., Cheng et al., 1988, *J. Biol. Chem.* 263: 15110; Ferrin and Camerini-Otero, 1991, *Science* 354:1494; Ramdas et al., 1989, *J. Biol. Chem.* 264:17395; Strobel et al., 1991, *Science* 254:1639; and Rigas et al., 1986, *Proc. Natl. Acad. Sci. U.S.A.* 83: 9591; each of which is incorporated herein by reference) and the hTERT mRNA and/or gene sequence. Typically, the triplex-forming oligonucleotides of the invention comprise a specific sequence of from about 10 to at least about 25 nucleotides or longer "complementary" to a specific sequence in the hTERT RNA or gene (i.e., large enough to form a stable triple helix, but small enough, depending on the mode of delivery, to administer *in vivo*, if desired). In this context, "complementary" means able to form a stable triple helix. In one embodiment, oligonucleotides are designed to bind specifically to the regulatory regions of the hTERT gene (e.g., the hTERT 5'-flanking sequence, promoters, and enhancers) or to the transcription initiation site, (e.g., between -10 and +10 from the transcription initiation site). For a review of recent therapeutic advances using triplex DNA, see Gee et al., in Huber and Carr, 1994, *MOLECULAR AND IMMUNOLOGIC APPROACHES*, Futura Publishing Co, Mt Kisco NY and Rininsland et al., 1997, *Proc. Natl. Acad. Sci. USA* 94:5854, which are both incorporated herein by reference.

c) ADMINISTRATION OF OLIGONUCLEOTIDES

Typically, the therapeutic methods of the invention involve the administration of an oligonucleotide that functions to inhibit or stimulate telomerase

D) 5'-GGACACCTGGCGGAAGGAGGG

can be used to inhibit telomerase activity. At 10 micromolar concentration each oligonucleotide, mixtures of oligonucleotides A and B; A, B, C, and D; and A, C, and D
5 inhibited telomerase activity in 293 cells when treated once per day for seven days. Inhibition was also observed when an antisense hTR molecule (5'-GCTCTAGAATGAAGGGTG-3') was used in combination with oligonucleotides A, B, and C; A, B, and D; and A and C. Useful control oligonucleotides in such experiments include:

- 10 S1) 5'-GCGACGACTGACATTGGCCGG
S2) 5'-GGCTCGAAGTAGCACCGGTGC
S3) 5'-GTGGGAACAGGCCGATGTCCC

To determine the optimum antisense oligonucleotide of the invention for
15 the particular application of interest, one can perform a scan using antisense oligonucleotide sets of the invention. One illustrative set is the set of 30-mer oligonucleotides that span the hTRT mRNA and are offset one from the next by fifteen nucleotides (i.e., ON1 corresponds to positions 1-30 and is TCCCACGTGCGCAGCAGGACGCGCTGC, ON2 corresponds to positions
20 16-45 and is GCCGGGGCCAGGGCTTCCCACGTGCGCAGC, and ON3 corresponds to positions 31-60 and is GGCATCGCGGGGTGGCCGGGGCCAGGGCT, and so on to the end of the mRNA). Each member of this set can be tested for inhibitory activity as disclosed herein. Those oligonucleotides that show inhibitory activity under the conditions of interest then identify a region of interest, and other oligonucleotides of the
25 invention corresponding to the region of interest (i.e., 8-mers, 10-mers, 15-mers, and so on) can be tested to identify the oligonucleotide with the preferred activity for the application.

For general methods relating to antisense polynucleotides, see ANTISENSE RNA AND DNA, (1988), D.A. Melton, Ed., Cold Spring Harbor Laboratory, Cold
30 Spring Harbor, NY). See also, Dagle et al., 1991, *Nucleic Acids Research*, 19:1805. For a review of antisense therapy, see, e.g., Uhlmann et al., *Chem. Reviews*, 90:543-584 (1990).

hTERT sequence (e.g., gene replacement and "gene knockout," respectively). Numerous other embodiments will be evident to one of skill upon review of the disclosure herein. In one embodiment, a vector encoding hTERT is also introduced. In another embodiment, vectors encoding telomerase-associated proteins are also introduced with or without a
5 vector for hTERT.

Vectors useful in hTERT gene therapy can be viral or nonviral, and include those described *supra* in relation to the hTERT expression systems of the invention. It will be understood by those of skill in the art that gene therapy vectors may comprise promoters and other regulatory or processing sequences, such as are described in this
10 disclosure. Usually the vector will comprise a promoter and, optionally, an enhancer (separate from any contained within the promoter sequences) that serve to drive transcription of an oligoribonucleotide, as well as other regulatory elements that provide for episomal maintenance or chromosomal integration and for high-level transcription, if desired. A plasmid useful for gene therapy can comprise other functional elements, such
15 as selectable markers, identification regions, and other sequences. The additional sequences can have roles in conferring stability both outside and within a cell, targeting delivery of hTERT nucleotide sequences (sense or antisense) to a specified organ, tissue, or cell population, mediating entry into a cell, mediating entry into the nucleus of a cell and/or mediating integration within nuclear DNA. For example, aptamer-like DNA
20 structures, or other protein binding moieties sites can be used to mediate binding of a vector to cell surface receptors or to serum proteins that bind to a receptor thereby increasing the efficiency of DNA transfer into the cell. Other DNA sites and structures can directly or indirectly bind to receptors in the nuclear membrane or to other proteins that go into the nucleus, thereby facilitating nuclear uptake of a vector. Other DNA
25 sequences can directly or indirectly affect the efficiency of integration.

Suitable gene therapy vectors may, or may not, have an origin of replication. For example, it is useful to include an origin of replication in a vector for propagation of the vector prior to administration to a patient. However, the origin of replication can often be removed before administration if the vector is designed to
30 integrate into host chromosomal DNA or bind to host mRNA or DNA. In some situations (e.g., tumor cells) it may not be necessary for the exogenous DNA to

activity under *in vivo* physiological conditions, and is relatively stable under those conditions for a period of time sufficient for a therapeutic effect. Modified nucleic acids may be useful in imparting such stability, as well as for targeting delivery of the oligonucleotide to the desired tissue, organ, or cell.

5 Oligo- and poly-nucleotides can be delivered directly as a drug in a suitable pharmaceutical formulation, or indirectly by means of introducing a nucleic acid into a cell, including liposomes, immunoliposomes, ballistics, direct uptake into cells, and the like as described herein. For treatment of disease, the oligonucleotides of the invention will be administered to a patient in a therapeutically effective amount. A
10 therapeutically effective amount is an amount sufficient to ameliorate the symptoms of the disease or modulate telomerase activity in the target cell, e.g., as can be measured using a TRAP assay or other suitable assay of telomerase biological function. Methods useful for delivery of oligonucleotides for therapeutic purposes are described in U.S. Patent 5,272,065, incorporated herein by reference. Other details of administration of
15 pharmaceutically active compounds are provided below. In another embodiment, oligo- and poly-nucleotides can be delivered using gene therapy and recombinant DNA expression plasmids of the invention.

3) GENE THERAPY

20 Gene therapy refers to the introduction of an otherwise exogenous polynucleotide which produces a medically useful phenotypic effect upon the (typically) mammalian cell(s) into which it is transferred. In one aspect, the present invention provides gene therapy methods and compositions for treatment of telomerase-associated conditions. In illustrative embodiments, gene therapy involves introducing into a cell a
25 vector that expresses an hTERT gene product (such as an hTERT protein substantially similar to the hTERT polypeptide having a sequence of SEQUENCE ID NO: 2, e.g., to increase telomerase activity, or an inhibitory hTERT polypeptide to reduce activity), expresses a nucleic acid having an hTERT gene or mRNA sequence (such as an antisense RNA, e.g., to reduce telomerase activity), expresses a polypeptide or polynucleotide
30 that otherwise affects expression of hTERT gene products (e.g., a ribozyme directed to hTERT mRNA to reduce telomerase activity), or replaces or disrupts an endogenous

expression of the protein encoded by the reporter gene. If, instead of a reporter protein, the encoded protein is toxic to the cell, activation of the promoter leads to cell morbidity or death. In one embodiment of the present invention, a vector comprising an hTRT promoter operably linked to a gene encoding a toxic protein is introduced into cells, such as human cells, e.g., cells in a human patient, resulting in cell death of cells in which hTRT promoter activating factors are expressed, such as cancer cells. In a related embodiment, the encoded protein is not itself toxic to a cell, but encodes an activity that renders the cell sensitive to an otherwise nontoxic drug. For example, tumors can be treated by introducing an hTRT-promoter-Herpes thymidine kinase (TK) gene fusion construct into tumor cells, and administering gancyclovir or the equivalent (see, e.g., Moolton and Wells, 1990, *J. Nat'l Canc. Inst.* 82:297). The art knows of numerous other suitable toxic or potentially toxic proteins and systems (using promoter sequences other than hTRT) that may be modified and applied in accordance with the present invention by one of skill in the art upon review of this disclosure.

Gene therapy vectors may be introduced into cells or tissues *in vivo*, *in vitro* or *ex vivo*. For *ex vivo* therapy, vectors may be introduced into cells, e.g., stem cells, taken from the patient and clonally propagated for autologous transplant back into the same patient (see, e.g., U.S. Patent Nos. 5,399,493 and 5,437,994, the disclosures of which are herein incorporated by reference). Cells that can be targeted for hTRT gene therapy aimed at increasing the telomerase activity of a target cell include, but are not limited to, embryonic stem or germ cells, particularly primate or human cells, as noted *supra*, hematopoietic stem cells (AIDS and post-chemotherapy), vascular endothelial cells (cardiac and cerebral vascular disease), skin fibroblasts and basal skin keratinocytes (wound healing and burns), chondrocytes (arthritis), brain astrocytes and microglial cells (Alzheimer's Disease), osteoblasts (osteoporosis), retinal cells (eye diseases), and pancreatic islet cells (Type I diabetes) and any of the cells listed in Table 3, *infra*, as well as any other cell types known to divide.

F) PHARMACEUTICAL COMPOSITIONS

In related aspects, the invention provides pharmaceutical compositions that comprise hTRT oligo- and poly-nucleotides, agonists, antagonists, or inhibitors,

integrate stably into the transduced cell, because transient expression may suffice to kill the tumor cells.

As noted, the present invention also provides methods and reagents for gene replacement therapy (i.e., replacement by homologous recombination of an endogenous hTERT gene with a recombinant gene). Vectors specifically designed for
5 integration by homologous recombination may be used. Important factors for optimizing homologous recombination include the degree of sequence identity and length of homology to chromosomal sequences. The specific sequence mediating homologous recombination is also important, because integration occurs much more
10 easily in transcriptionally active DNA. Methods and materials for constructing homologous targeting constructs are described by e.g., Mansour et al., 1988, *Nature* 336: 348; Bradley et al., 1992, *Bio/Technology* 10: 534. See also, U.S. Patent Nos. 5,627,059; 5,487,992; 5,631,153; and 5,464,764. In one embodiment, gene replacement therapy involves altering or replacing all or a portion of the regulatory sequences
15 controlling expression of the hTERT gene that is to be regulated. For example, the hTERT promoter sequences (e.g., such as are found in SEQUENCE ID NO: 6) may be disrupted (to decrease hTERT expression or to abolish a transcriptional control site) or an exogenous promoter (e.g., to increase hTERT expression) substituted.

The invention also provides methods and reagents for hTERT "gene
20 knockout" (i.e., deletion or disruption by homologous recombination of an endogenous hTERT gene using a recombinantly produced vector). In gene knockout, the targeted sequences can be regulatory sequences (e.g., the hTERT promoter), or RNA or protein coding sequences. The use of homologous recombination to alter expression of endogenous genes is described in detail in U.S. Patent No. 5,272,071 (and the U.S.
25 Patents cited *supra*), WO 91/09955, WO 93/09222, WO 96/29411, WO 95/31560, and WO 91/12650. See also, Moynahan et al., 1996, *Hum. Mol. Genet.* 5:875.

The invention further provides methods for specifically killing telomerase-positive cells, or preventing transformation of telomerase negative cells to a telomerase positive state, using the hTERT gene promoter to regulate expression of a
30 protein toxic to the cell. As shown in Example 14, an hTERT promoter sequence may be operably linked to a reporter gene such that activation of the promoter results in

carboxymethylcellulose; and gums including arabic and tragacanth; as well as proteins such as gelatin and collagen. If desired, disintegrating or solubilizing agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, alginic acid, or a salt thereof, such as sodium alginate.

5 Dragee cores are provided with suitable coatings such as concentrated sugar solutions, which may also contain gum arabic, talc, polyvinylpyrrolidone, carbopol (RTM) gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for product identification or to characterize the quantity of active
10 compound (*i.e.*, dosage).

 Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with a filler or binders such as lactose or starches, lubricants such as talc or magnesium
15 stearate, and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid paraffin, or liquid polyethylene glycol with or without stabilizers.

 Pharmaceutical formulations for parenteral administration include aqueous solutions of active compounds. For injection, the pharmaceutical compositions
20 of the invention may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hank's solution, Ringer's solution, or physiologically buffered saline. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Additionally, suspensions of the active compounds may be prepared as
25 appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate or triglycerides, or liposomes. Optionally, the suspension may also contain suitable stabilizers or agents which increase the solubility of the compounds to allow for the preparation of highly concentrated solutions.

30 For topical or nasal administration, penetrants appropriate to the particular barrier to be permeated are used in the formulation. Such penetrants are

alone or in combination with at least one other agent, such as a stabilizing compound, diluent, carrier, or another active ingredient or agent.

The therapeutic agents of the invention may be administered in any sterile, biocompatible pharmaceutical carrier, including, but not limited to, saline, buffered saline, dextrose, and water. Any of these molecules can be administered to a patient alone, or in combination with other agents, drugs or hormones, in pharmaceutical compositions where it is mixed with suitable excipient(s), adjuvants, and/or pharmaceutically acceptable carriers. In one embodiment of the present invention, the pharmaceutically acceptable carrier is pharmaceutically inert.

Administration of pharmaceutical compositions is accomplished orally or parenterally. Methods of parenteral delivery include topical, intra-arterial (e.g., directly to the tumor), intramuscular, subcutaneous, intramedullary, intrathecal, intraventricular, intravenous, intraperitoneal, or intranasal administration. In addition to the active ingredients, these pharmaceutical compositions may contain suitable pharmaceutically acceptable carriers comprising excipients and other compounds that facilitate processing of the active compounds into preparations which can be used pharmaceutically. Further details on techniques for formulation and administration may be found in the latest edition of "REMINGTON'S PHARMACEUTICAL SCIENCES" (Maack Publishing Co, Easton PA).

Pharmaceutical compositions for oral administration can be formulated using pharmaceutically acceptable carriers well known in the art in dosages suitable for oral administration. Such carriers enable the pharmaceutical compositions to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, etc., suitable for ingestion by the patient. See PCT publication WO 93/23572.

Pharmaceutical preparations for oral use can be obtained through combination of active compounds with solid excipient, optionally grinding a resulting mixture, and processing the mixture of granules, after adding suitable additional compounds, if desired, to obtain tablets or dragee cores. Suitable excipients are carbohydrate or protein fillers include, but are not limited to sugars, including lactose, sucrose, mannitol, or sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose such as methyl cellulose, hydroxypropylmethyl-cellulose, or sodium

model is also used to achieve a desirable concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective amount refers to that amount of oligo- or polynucleotide, agonist or antagonists which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals (*e.g.*, ED₅₀, the dose therapeutically effective in 50% of the population; and LD₅₀, the dose lethal to 50% of the population). The dose ratio between therapeutic and toxic effects is the therapeutic index, and it can be expressed as the ratio, ED₅₀/LD₅₀. Pharmaceutical compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies is used in formulating a range of dosage for human use. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED₅₀ with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, sensitivity of the patient, and the route of administration.

The exact dosage is chosen by the individual physician in view of the patient to be treated. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Additional factors which may be taken into account include the severity of the disease state (*e.g.*, tumor size and location; age, weight and gender of the patient; diet, time and frequency of administration, drug combination(s), reaction sensitivities, and tolerance/response to therapy). Long acting pharmaceutical compositions might be administered every 3 to 4 days, every week, or once every two weeks depending on half-life and clearance rate of the particular formulation. Guidance as to particular dosages and methods of delivery is provided in the literature (see, US Patent Nos. 4,657,760; 5,206,344; and 5,225,212, herein incorporated by reference). Those skilled in the art will typically employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides can be specific to particular cells, conditions, locations, and the like.

generally known in the art.

The pharmaceutical compositions of the present invention may be manufactured in a manner similar to that known in the art (*e.g.*, by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping or lyophilizing processes).

The pharmaceutical composition may be provided as a salt and can be formed with many acids, including but not limited to hydrochloric, sulfuric, acetic, lactic, tartaric, malic, succinic, etc. Salts tend to be more soluble in aqueous or other protonic solvents that are the corresponding free base forms. In other cases, the preferred preparation may be a lyophilized powder in 1 mM-50 mM histidine, 0.1%-2% sucrose, 2%-7% mannitol at a pH range of 4.5 to 5.5, that is combined with buffer prior to use.

After pharmaceutical compositions comprising a compound of the invention formulated in a acceptable carrier have been prepared, they can be placed in an appropriate container and labeled for treatment of an indicated condition. For administration of human telomerase nucleic acids, such labeling would include amount, frequency and method of administration.

Pharmaceutical compositions suitable for use in the present invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. "Therapeutically effective amount" or "pharmacologically effective amount" are well recognized phrases and refer to that amount of an agent effective to produce the intended pharmacological result. Thus, a therapeutically effective amount is an amount sufficient to ameliorate the symptoms of the disease being treated. One useful assay in ascertaining an effective amount for a given application (*e.g.*, a therapeutically effective amount) is measuring the effect on telomerase activity in a target cell. The amount actually administered will be dependent upon the individual to which treatment is to be applied, and will preferably be an optimized amount such that the desired effect is achieved without significant side-effects. The determination of a therapeutically effective dose is well within the capability of those skilled in the art.

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays or in any appropriate animal model. The animal

replicative capacity, of the invention are made by increasing telomerase activity in the cell. Any method disclosed herein for increasing telomerase activity can be used. Thus, in one embodiment, cells are immortalized by increasing the amount of an hTERT polypeptide in the cell. In one embodiment, hTERT levels are increased by introducing an hTERT expression vector into the cell (with stable transfection sometimes preferred). As discussed above, the hTERT coding sequence is usually operably linked to a promoter, which may be inducible or constitutively active in the cell.

In one embodiment, a polynucleotide comprising a sequence encoding a polypeptide of SEQUENCE ID NO: 2, which sequence is operably linked to a promoter (e.g., a constitutively expressed promoter, e.g., a sequence of SEQUENCE ID NO: 6), is introduced into the cell. In one embodiment the polynucleotide comprises a sequence of SEQUENCE ID NO: 1. Preferably the polynucleotide includes polyadenylation and termination signals. In other embodiments, additional elements such as enhancers or others discussed *supra* are included. In an alternative embodiment, the polynucleotide does not include a promoter sequence, such sequence being provided by the target cell endogenous genome following integration (e.g., recombination, e.g., homologous recombination) of the introduced polynucleotide. The polynucleotide may be introduced into the target cell by any method, including any method disclosed herein, such as lipofection, electroporation, virosomes, liposomes, immunoliposomes, polycation:nucleic acid conjugates, naked DNA).

Using the methods of the invention, any vertebrate cell can be caused to have an increased proliferative capacity or even be immortalized and sustained indefinitely in culture. In one embodiment the cells are mammalian, with human cells preferred for many applications. Examples of human cells that can be immortalized include those listed in Table 3A.

It will be recognized that the "diagnostic" assays of the invention described *infra* may be used to identify and characterize the immortalized cells of the invention.

VIII) INCREASING PROLIFERATIVE CAPACITY AND PRODUCTION OF IMMORTALIZED CELLS, CELL LINES, AND ANIMALS

As discussed above, most vertebrate cells senesce after a finite number of divisions in culture (e.g., 50 to 100 divisions). Certain variant cells, however, are able to divide indefinitely in culture (e.g., HeLa cells, 293 cells) and, for this reason, are useful for research and industrial applications. Usually these immortal cell lines are derived from spontaneously arising tumors, or by transformation by exposure to radiation or a tumor-inducing virus or chemical. Unfortunately, a limited selection of cell lines, especially human cell lines representing differentiated cell function, is available. Moreover, the immortal cell lines presently available are characterized by chromosomal abnormalities (e.g., aneuploidy, gene rearrangements, or mutations). Further, many long-established cell lines are relatively undifferentiated (e.g., they do not produce highly specialized products of the sort that uniquely characterize particular tissues or organs). Thus, there is a need for new methods of generating immortal cells, especially human cells. One use for immortalized cells is in production of natural proteins and recombinant proteins (e.g., therapeutic polypeptides such as erythropoietin, human growth hormone, insulin, and the like), or antibodies, for which a stable, genetically normal cell line is preferred. For production of some recombinant proteins, specialized cell types may also be preferred (e.g., pancreatic cells for the production of human insulin). Another use for immortalized cells or even mortal cells with increased proliferative capacity (relative to unmodified cells) is for introduction into a patient for gene therapy, or for replacement of diseased or damaged cells or tissue. For example, autologous immune cells containing or expressing a, e.g., recombinant hTERT gene or polypeptide of the invention can be used for cell replacement in a patient after aggressive cancer therapy, e.g., whole body irradiation. Another use for immortalized cells is for *ex vivo* production of "artificial" tissues or organs (e.g., skin) for therapeutic use. Another use for such cells is for screening or validation of drugs, such as telomerase-inhibiting drugs, or for use in production of vaccines or biological reagents. Additional uses of the cells of the invention will be apparent to those of skill.

The immortalized cells and cell lines, as well as those of merely increased

- zymogenic cell of gastric gland, secreting pepsinogen
oxyntic cell of gastric gland, secreting HCl
acinar cell of pancreas, secreting digestive enzymes and bicarbonate
- 5 Paneth cell of small intestine, secreting lysozyme
type II pneumocyte of lung, secreting surfactant
Clara cell of lung
- Cells specialized for Secretion of Hormones**
- 10 cells of anterior pituitary, secreting
growth hormone, follicle-stimulating hormone,
luteinizing hormone, prolactin, adrenocorticotrophic hormone, and thyroid-stimulating hormone,
cell of intermediate pituitary, secreting
15 melanocyte-stimulating hormone
cells of posterior pituitary, secreting
oxytocin, vasopressin
cells of gut, secreting
serotonin, endorphin, somatostatin, gastrin,
20 secretin, cholecystokinin, insulin and glucagon
cells of thyroid gland, secreting
thyroid hormone, calcitonin
cells of parathyroid gland, secreting
parathyroid hormone, oxyphil cell
25 cells of adrenal gland, secreting
epinephrine, norepinephrine, and steroid hormones;
mineralocorticoids
glucocorticoids
cells of gonads, secreting
30 testosterone (Leydig cell of testis)
estrogen (theca interna cell of ovarian follicle)
progesterone (corpus luteum cell of ruptured ovarian follicle)
cells of juxtaglomerular apparatus of kidney
35 juxtaglomerular cell (secreting renin)
macula densa cell
peripolar cell
mesangial cell
- Epithelial Absorptive Cells in Gut, Exocrine Glands, and Urogenital Tract**
- 40 brush border cell of intestine (with microvilli)
striated duct cell of exocrine glands
gall bladder epithelial cell
45 brush border cell of proximal tubule of kidney
distal tubule cell of kidney
nonciliated cell of ductulus efferens
epididymal principal cell
epididymal basal cell
50
- Cells Specialized for Metabolism and Storage**
hepatocyte (liver cell)
fat cells
white fat
55 brown fat

TABLE 3A

HUMAN CELLS IN WHICH hTRT EXPRESSION MAY BE INCREASED

Keratinizing Epithelial Cells

- keratinocyte of epidermis (differentiating epidermal cell)
- 5 basal cell of epidermis (stem cell)
- keratinocyte of fingernails and toenails
- basal cell of nail bed (stem cell)
- hair shaft cells
- 10 medullary, cortical, cuticular; hair-root sheath cells,
- cuticular, of Huxley's layer, of Henle's layer
- external; hair matrix cell (stem cell)

Cells of Wet Stratified Barrier Epithelia

- surface epithelial cell of stratified squamous epithelium of
- 15 tongue, oral cavity, esophagus, anal canal, distal urethra,
- vagina
- basal cell of these epithelia (stem cell)
- cell of external corneal epithelium
- cell of urinary epithelium (lining bladder and urinary
- 20 ducts)

Epithelial Cells Specialized for Exocrine Secretion

- cells of salivary gland
- 25 mucous cell (secretion rich in polysaccharide)
- serous cell (secretion rich in glycoprotein enzymes)
- cell of von Ebner's gland in tongue (secretion to wash
- over taste buds)
- cell of mammary gland, secreting milk
- cell of lacrimal gland, secreting tears
- 30 cell of ceruminous gland of ear, secreting wax
- cell of eccrine sweat gland, secreting glycoproteins (dark
- cell)
- cell of eccrine sweat gland, secreting small molecules
- (clear cell)
- 35 cell of apocrine sweat gland (odoriferous secretion,
- sex-hormone sensitive)
- cell of gland of Moll in eyelid (specialized sweat gland)
- cell of sebaceous gland, secreting lipid-rich sebum
- cell of Bowman's gland in nose (secretion to wash over
- 40 olfactory epithelium)
- cell of Brunner's gland in duodenum, secreting alkaline
- solution of mucus and enzymes
- cell of seminal vesicle, secreting components of seminal
- fluid, including fructose (as fuel for swimming sperm)
- 45 cell of prostate gland, secreting other components of
- seminal fluid
- cell of bulbourethral gland, secreting mucus
- cell of Bartholin's gland, secreting vaginal lubricant
- cell of gland of Littré, secreting mucus
- 50 cell of endometrium of uterus, secreting mainly
- carbohydrates
- isolated goblet cell of respiratory and digestive tracts,
- secreting mucus
- mucous cell of lining of stomach

- planum semilunatum cell of vestibular apparatus of ear
 (secreting proteoglycan)
 interdental cell of organ of Corti (secreting tectorial
 "membrane" covering hair cells of organ of Corti)
- 5 nonepithelial (connective tissue)
 fibroblasts (various-of loose connective tissue, of
 cornea, of tendon, of reticular tissue of bone marrow, etc.)
 pericyte of blood capillary
 nucleus pulposus cell of intervertebral disc
- 10 cementoblast/cementocyte (secreting bonelike cementum of
 root of tooth)
 odontoblast/odontocyte (secreting dentin of tooth)
 chondrocytes
 of hyaline cartilage, of fibrocartilage, of elastic
- 15 cartilage
 osteoblast/osteocyte
 osteoprogenitor cell (stem cell of osteoblasts)
 hyalocyte of vitreous body of eye
 stellate cell of perilymphatic space of ear
- 20 **Contractile Cells**
 skeletal muscle cells
 red (slow)
 white (fast)
- 25 intermediate
 muscle spindle - nuclear bag
 muscle spindle - nuclear chain
 satellite cell (stem cell)
- heart muscle cells
 30 ordinary
 nodal
 Purkinje fiber
 smooth muscle cells
 myoepithelial cells
- 35 of iris
 of exocrine glands
- Cells of Blood and Immune System**
 red blood cell
- 40 megakaryocyte
 macrophages
 monocyte
 connective tissue macrophage (various)
 Langerhans cell (in epidermis)
- 45 osteoclast (in bone)
 dendritic cell (in lymphoid tissues)
 microglial cell (in central nervous system)
- neutrophil
 eosinophil
- 50 basophil
 mast cell
 T lymphocyte
 helper T cell
 suppressor T cell
- 55 killer T cell

lipocyte of liver

- 5 **Epithelial Cells Serving Primarily a Barrier Function,**
 Lining the Lung, Gut, Exocrine Glands, and Urogenital Tract
 type I pneumocyte (lining air space of lung)
 pancreatic duct cell (centroacinar cell)
10 nonstriated duct cell of sweat gland, salivary gland,
 mammary gland
 parietal cell of kidney glomerulus
 podocyte of kidney glomerulus
 cell of thin segment of loop of Henle (in kidney)
15 collecting duct cell (in kidney)
 duct cell of seminal vesicle, prostate gland

- Epithelial Cells Lining Closed Internal Body Cavities**
 vascular endothelial cells of blood vessels and lymphatics
20 fenestrated
 continuous
 splenic
 synovial cell (lining joint cavities, secreting largely
 hyaluronic acid)
25 serosal cell (lining peritoneal, pleural, and pericardial
 cavities)
 squamous cell lining perilymphatic space of ear
 cells lining endolymphatic space of ear
 squamous cell
30 columnar cells of endolymphatic sac
 with microvilli
 without microvilli
 "dark" cell
 vestibular membrane cell (resembling choroid plexus cell)
35 stria vascularis basal cell
 stria vascularis marginal cell
 cell of Claudius
 cell of Boettcher
 choroid plexus cell (secreting cerebrospinal fluid)
40 squamous cell of pia-arachnoid
 cells of ciliary epithelium of eye
 pigmented
 nonpigmented
 corneal "endothelial" cell

- 45 **Ciliated Cells with Propulsive Function**
 of respiratory tract
 of oviduct and of endometrium of uterus (in female)
 of rete testis and ductulus efferens (in male)
50 of central nervous system (ependymal cell lining brain
 cavities)

- Cells Specialized for Secretion of Extracellular Matrix**
 epithelial:
55 ameloblast (secreting enamel of tooth)

- Hensen cell
- supporting cell of vestibular apparatus
- supporting cell of taste bud (type I taste bud cell)
- supporting cell of olfactory epithelium
- 5 Schwann cell
- satellite cell (encapsulating peripheral nerve cell bodies)
- enteric glial cell

- Neurons and Glial Cells of Central Nervous System**
- 10 neurons
- glial cells
 - astrocyte
 - oligodendrocyte

- 15 **Lens Cells**
- anterior lens epithelial cell
- lens fiber (crystallin-containing cell)

- Pigment Cells**
- 20 melanocyte , retinal pigmented epithelial cell

- Germ Cells**
- oogonium/oocyte
- spermatocyte
- 25 spermatogonium (stem cell for spermatocyte)

- Nurse Cells**
- ovarian follicle cell
- Sertoli cell (in testis)
- 30 thymus epithelial cell

- Stem Cells**
- embryonic stem cell
- embryonic germ cell
- 35 adult stem cell
- fetal stem cell

VIII) DIAGNOSTIC ASSAYS

A) INTRODUCTION

40 1) TRT ASSAYS

The present invention provides a wide variety of assays for TRT, preferably hTRT, and telomerase. These assays provide, *inter alia*, the basis for sensitive, inexpensive, convenient, and widely applicable assays for diagnosis and prognosis of a number of human diseases, of which cancer is an illustrative example.

- 45 As noted *supra*, hTRT gene products (protein and mRNA) are usually elevated in immortal human cells relative to most normal mortal cells (i.e., telomerase-negative cells and most telomerase-positive normal adult somatic cells). Thus, in one aspect, the invention provides assays useful for detecting or measuring the presence, absence,

- B lymphocyte
 IgM
 IgG
 IgA
 5 IgE
 killer cell
 stem cells for the blood and immune system (various)
- Sensory Transducers**
 10 photoreceptors
 rod
 cones
 blue sensitive
 green sensitive
 15 red sensitive
 hearing
 inner hair cell of organ of Corti
 outer hair cell of organ of Corti
 acceleration and gravity
 20 type I hair cell of vestibular apparatus of ear
 type II hair cell of vestibular apparatus of ear
 taste
 type II taste bud cell
 smell
 25 olfactory neuron
 basal cell of olfactory epithelium (stem cell for
 olfactory
 neurons)
 blood Ph
 30 carotid body cell
 type I
 type II
 touch
 Merkel cell of epidermis
 35 primary sensory neurons specialized for touch temperature
 primary sensory neurons specialized for temperature
 cold sensitive
 heat sensitive
 pain
 40 primary sensory neurons specialized for pain
 configurations and forces in musculoskeletal system
 proprioceptive primary sensory neurons
- Autonomic Neurons**
 45 cholinergic
 adrenergic
 peptidergic
- Supporting Cells of Sense Organs and of Peripheral Neurons**
 50 supporting cells of organ of Corti
 inner pillar cell
 outer pillar cell
 inner phalangeal cell
 outer phalangeal cell
 55 border cell

infertile, activated), or status (e.g., fertile), and the term "prognostic" has its usual meaning of predicting the probable development and/or outcome of a disease or condition. Although these two terms are used in somewhat different ways in a clinical setting, it will be understood that any of the assays or assay formats disclosed
5 below in reference to "diagnosis" are equally suitable for determination of prognosis because it is well established that higher telomerase activity levels are associated with poorer prognoses for cancer patients, and because the present invention provides detection methods specific for hTERT, which is expressed at levels that closely correlate with telomerase activity in a cell .

10

2) DIAGNOSIS AND PROGNOSIS OF CANCER

The assays described below are useful for detecting certain variations in hTERT gene sequence (mutations and heritable hTERT alleles) that are indicative of a predilection for cancers or other conditions associated with abnormal regulation of
15 telomerase activity (infertility, premature aging).

3) MONITORING CELLS IN CULTURE

The assays described herein are also useful for characterization of hTERT genes in cells *ex vivo* or *in vitro*. Because elevated hTERT levels are
20 characteristic of immortalized cells, the assays of the invention can be used, for example, to screen for, or identify, immortalized cells or to identify an agent capable of mortalizing immortalized cells by inhibiting hTERT expression or function. For example, the assay will be useful for identifying cells immortalized by increased expression of hTERT in the cell, e.g., by the expression of a recombinant hTERT or by
25 increased expression of an endogenously coded hTERT (e.g., by promoter activation).

Similarly, these assays may be used to monitor hTERT expression in transgenic animals or cells (e.g., yeast or human cells containing an hTERT gene). In particular, the effects of certain treatments (e.g., application of known or putative telomerase ant agonists) on the hTERT levels in human and nonhuman cells expressing
30 the hTERT of the invention can be used for identifying useful drugs and drug candidates (e.g., telomerase activity-modulating drugs).

or quantity of an hTERT gene product in a sample from, or containing, human or other mammalian or eukaryotic cells to characterize the cells as immortal (such as a malignant tumor cell) or mortal (such as most normal somatic cells in adults) or as telomerase positive or negative.

5 Any condition characterized by the presence or absence of an hTERT gene product (i.e., protein or RNA) may be diagnosed using the methods and materials described herein. These include, as described more fully below, cancers, other diseases of accelerated cell proliferation, immunological disorders, fertility, infertility, and others. Moreover, because the degree to which telomerase activity is
10 elevated in cancer cells is correlated with characteristics of the tumor, such as metastatic potential, monitoring hTERT, mRNA or protein levels can be used to estimate and predict the likely future progression of a tumor.

 In one aspect, host (e.g., patient) cells are assayed to identify nucleic acids with sequences characteristic of a heritable propensity for abnormal hTERT gene
15 expression (abnormal quantity, regulation, or product), such as is useful in genetic screening or genetic counseling. The methods described below in some detail are indicative of useful assays that can be carried out using the sequences and relationships disclosed herein. However, numerous variations or other applications of these assays will be apparent to those of ordinary skill in the art in view of this
20 disclosure.

 It will be recognized that, although the assays below are presented in terms of diagnostic and prognostic methods, they may be used whenever an hTERT gene, gene product, or variant is to be detected, quantified, or characterized. Thus, for example, the "diagnostic" methods described *infra* are useful for assays of hTERT
25 or telomerase during production and purification of hTERT or human telomerase, for characterization of cell lines derived from human cells (e.g., to identify immortal lines), for characterization of cells, non-human animals, plants, fungi, bacteria or other organisms that comprise a human TRT gene or gene product (or fragments thereof).

30 As used herein, the term "diagnostic" has its usual meaning of identifying the presence or nature of a disease (e.g., cancer), condition (e.g.,

Tris-buffer, or the like, at physiological pH can be used.

A "biological sample" obtained from a patient can be referred to either as a "biological sample" or a "patient sample." It will be appreciated that analysis of a "patient sample" need not necessarily require removal of cells or tissue from the patient. For example, appropriately labeled hTERT-binding agents (e.g., antibodies or nucleic acids) can be injected into a patient and visualized (when bound to the target) using standard imaging technology (e.g., CAT, NMR, and the like.)

C) NUCLEIC ACID ASSAYS

In one embodiment, this invention provides methods for detecting and analyzing normal or abnormal hTERT genes (or fragments thereof). The form of such qualitative or quantitative assays may include, but is not limited to, amplification-based assays with or without signal amplification, hybridization based assays, and combination amplification-hybridization assays. It will be appreciated by those of skill that the distinction between hybridization and amplification is for convenience only: as illustrated in the examples below, many assay formats involve elements of both hybridization and amplification, so that the categorization is somewhat arbitrary in some cases.

1) PREPARATION OF NUCLEIC ACIDS

In some embodiments, nucleic acid assays are performed with a sample of nucleic acid isolated from the cell, tissue, organism, or cell line to be tested. The nucleic acid may be "isolated" from the sample according to any of a number of methods well known to those of skill in the art. In this context, "isolated" refers to any separation of the species or target to be detected from any other substance in the mixture, but does not necessarily indicate a significant degree of purification of the target. One of skill will appreciate that, where alterations in the copy number of the hTERT gene are to be detected, genomic DNA is the target to be detected. Methods for isolating nucleic acids are well known to those of skill in the art and are described, for example, Tijssen, P. ed. of LABORATORY TECHNIQUES IN BIOCHEMISTRY AND MOLECULAR BIOLOGY: HYBRIDIZATION WITH NUCLEIC ACID

PROBES, PART I. THEORY AND NUCLEIC ACID PREPARATION, Elsevier, N.Y. (1993)
Chapt. 3, which is incorporated herein by reference. In one embodiment, the total
nucleic acid is isolated from a given sample using an acid guanidinium-phenol-
chloroform extraction method and poly(A)+ mRNA is isolated by oligo-dT column
5 chromatography or by using (dT)n magnetic beads (*see, e.g.*, Sambrook et al., and
Ausubel et al., *supra*).

In alternative embodiments, it is not necessary to isolate nucleic acids
from the biological sample prior to carrying out amplification, hybridization or other
assays. For example, many amplification techniques such as PCR defined above can
10 be carried out using permeabilized cells (histological specimens and FACS analyses),
whole lysed cells, or crude cell fractions such as certain cell extracts. Preferably,
steps are taken to preserve the integrity of the target nucleic acid if necessary.
Amplification and hybridization assays can also be carried out *in situ*, for example, in
thin tissue sections from a biopsy sample or from a cell monolayer (e.g., blood cells
15 or disaggregated tissue culture cells). Amplification can also be carried out in an
intact whole cell or fixed cells. For example, PCR or LCR amplification methods
may be carried out, as is well known in the art, *in situ*, e.g., using a polymerase or
ligase, a primer or primer(s), and (deoxy)ribonucleoside triphosphates (if a
polymerase is employed) on fixed, permeabilized, or microinjected cells to amplify
20 target hTERT DNA. Cells containing an hTERT DNA sequence of interest can then be
detected. This method is often useful when fluorescently-labeled dNTPs, primers, or
other components are used in conjunction with microscopy, FACS analysis or the
equivalent.

25 2) AMPLIFICATION BASED ASSAYS

In one embodiment, the assays of the present invention are
amplification-based assays for detection of an hTERT gene. In an amplification based
assay, all or part of an hTERT gene (hereinafter also referred to as "target") is
amplified, and the amplification product is then detected directly or indirectly. When
30 there is no underlying gene to act as a template, no amplification product is produced
(e.g., of the expected size), or amplification is non-specific and typically there is no

single amplification product. In contrast, when the underlying gene is present, the target sequence is amplified, providing an indication of the presence and/or quantity of the underlying gene. Target amplification-based assays are well known to those of skill in the art.

5 The present invention provides a wide variety of primers and probes for detecting hTRT genes. Such primers and probes are sufficiently complementary to the hTRT gene to hybridize to the target nucleic acid. Primers are typically at least 6 bases in length, usually between about 10 and about 100 bases, typically between about 12 and about 50 bases, and often between about 14 and about 25 bases
10 in length. One of skill, having reviewed the present disclosure, will be able, using routine methods, to select primers to amplify all, or any portion, of the hTRT gene or to distinguish between variant gene products, hTRT alleles, and the like. Table 2 lists illustrative primers useful for PCR amplification of the hTRT, or specific hTRT gene products or regions. As is known in the art, single oligomers (e.g., U.S. Pat.
15 No. 5,545,522), nested sets of oligomers, or even a degenerate pool of oligomers may be employed for amplification, e.g., as illustrated by the amplification of the *Tetrahymena* TRT cDNA as described *infra*.

 The invention provides a variety of methods for amplifying and detecting an hTRT gene including the polymerase chain reaction (including all
20 variants, e.g., the Sunrise Amplification System (Oncor, Inc, Gaithersburg MD); and numerous others known in the art). In one illustrative embodiment, PCR amplification is carried out in a 50 µl solution containing the nucleic acid sample, 100 µM in each dNTP (dATP, dCTP, dGTP and dTTP; Pharmacia LKB Biotechnology, NJ), the hTRT-specific PCR primer(s), 1 unit/ Taq polymerase
25 (Perkin Elmer, Norwalk CT), 1x PCR buffer (50 mM KCl, 10 mM Tris, pH 8.3 at room temperature, 1.5 mM MgCl₂, 0.01% gelatin) with the amplification run for about 30 cycles at 94° for 45 sec, 55° for 45 sec and 72° for 90 sec. However, as will be appreciated, numerous variations may be made to optimize the PCR amplification for any particular reaction.

30 Other suitable target amplification methods include the ligase chain reaction (LCR; e.g., Wu and Wallace, 1989, *Genomics* 4:560; Landegren *et al.*,

1988, *Science*, 241: 1077, Barany, 1991, *Proc. Natl. Acad. Sci. USA* 88:189 and Barringer *et al.*, 1990, *Gene*, 89: 117); strand displacement amplification (SDA; e.g., Walker *et al.*, 1992, *Proc. Natl. Acad. Sci. U.S.A.* 89:392-396); transcription amplification (e.g., Kwoh *et al.*, 1989, *Proc. Natl. Acad. Sci. USA*, 86: 1173); self-sustained sequence replication (3SR; e.g., Fahy *et al.*, 1992, *PCR Methods Appl.* 1:25, Guatelli *et al.*, 1990, *Proc. Nat. Acad. Sci. USA*, 87: 1874); the nucleic acid sequence based amplification (NASBA, Cangene, Mississauga, Ontario; e.g., Compton, 1991, *Nature* 350:91); the transcription-based amplification system (TAS); and the self-sustained sequence replication system (SSR). Each of the
10 aforementioned publications is incorporated herein by reference. One useful variant of PCR is PCR ELISA (e.g., Boehringer Mannheim Cat. No. 1 636 111) in which digoxigenin-dUTP is incorporated into the PCR product. The PCR reaction mixture is denatured and hybridized with a biotin-labeled oligonucleotide designed to anneal to an internal sequence of the PCR product. The hybridization products are
15 immobilized on streptavidin coated plates and detected using anti-digoxigenin antibodies. Examples of techniques sufficient to direct persons of skill through *in vitro* amplification methods are found in PCR TECHNOLOGY: PRINCIPLES AND APPLICATIONS FOR DNA AMPLIFICATION, H. Erlich, Ed. Freeman Press, New York, NY (1992); PCR PROTOCOLS: A GUIDE TO METHODS AND APPLICATIONS, eds. Innis, Gelfand, Snisky, and White, Academic Press, San Diego, CA (1990); Mattila *et al.*, 1991, *Nucleic Acids Res.* 19: 4967; Eckert and Kunkel, (1991) PCR METHODS AND APPLICATIONS 1: 17; PCR, eds. McPherson, Quirk, and Taylor, IRL Press, Oxford; U.S. Patent Nos. 4,683,195, 4,683,202, and 4,965,188; Barringer *et al.*, 1990, *Gene*, 89:117; Lomell *et al.*, 1989, *J. Clin. Chem.*, 35:1826, each of which is
25 incorporated herein for all purposes.

Amplified products may be directly analyzed, e.g., by size as determined by gel electrophoresis; by hybridization to a target nucleic acid immobilized on a solid support such as a bead, membrane, slide, or chip; by sequencing; immunologically, e.g., by PCR-ELISA, by detection of a fluorescent, phosphorescent, or radioactive signal; or by any of a variety of other well-known
30 means. For example, an illustrative example of a detection method uses PCR

primers augmented with hairpin loops linked to fluorescein and a benzoic acid derivative that serves as a quencher, such that fluorescence is emitted only when the primers unfold to bind their targets and replication occurs.

Methods for augmenting the ability to detect the amplified target
5 include signal amplification system such as: branched DNA signal amplification (e.g., U.S. Pat. No. 5,124,246; Urdea, 1994, *Bio/Tech.* 12:926); tyramide signal amplification (TSA) system (Du Pont); catalytic signal amplification (CSA; Dako); Q Beta Replicase systems (Tyagi et al., 1996, *Proc. Nat. Acad. Sci. USA*, 93: 5395); or the like.

10 One of skill in the art will appreciate that whatever amplification method is used, a variety of quantitative methods known in the art can be used if quantitation is desired. Co-amplification of the control polynucleotide (usually present at a known concentration or copy number) can be used for normalization to the cell number in the sample as compared to the amount of hTERT in the sample.
15 Suitable control polynucleotides for co-amplification reactions include DNA, RNA expressed from housekeeping genes, constitutively expressed genes, and *in vitro* synthesized RNAs or DNAs added to the reaction mixture. Endogenous control polynucleotides are those that are already present in the sample, while exogenous control polynucleotides are added to a sample, creating a "spiked" reaction.
20 Illustrative control RNAs include β -actin RNA, GAPDH RNA, snRNAs, hTERT, and endogenously expressed 28S rRNA (see Khan *et al.*, 1992, *Neurosci. Lett.* 147:114). Exogenous control polynucleotides include a synthetic AW106 cRNA, which may be synthesized as a sense strand from pAW106 by T7 polymerase. It will be appreciated that for the co-amplification method to be useful for quantitation, the control and
25 target polynucleotides must typically both be amplified in a linear range. Detailed protocols for quantitative PCR may be found in PCR PROTOCOLS, A GUIDE TO METHODS AND APPLICATIONS, Innis *et al.*, Academic Press, Inc. N.Y., (1990) and Ausubel *et al.*, *supra* (Unit 15) and Diaco, R. (1995) *Practical Considerations for the Design of Quantitative PCR Assays*, in PCR STRATEGIES, pg. 84-108, Innis *et al.*
30 eds, Academic Press, New York.

Depending on the sequence of the endogenous or exogenous standard,

different primer sets may be used for the co-amplification reaction. In one method, called competitive amplification, quantitative PCR involves simultaneously co-amplifying a known quantity of a control sequence using the same primers used for amplification of the target nucleic acid (one pair of 2 primers). In an alternative
5 embodiment, known as non-competitive competition, the control sequence and the target sequence are amplified using different primers (i.e., 2 pairs of 2 primers). In another alternative embodiment, called semi-competitive amplification, three primers are used, one of which is hTERT-specific, one of which is control specific, and one of which is capable of annealing to both the target and control sequences. Semi-
10 competitive amplification is described in U.S. Patent No. 5,629,154, which is incorporated herein by reference.

3) HYBRIDIZATION-BASED ASSAYS

a) GENERALLY

15 A variety of methods for specific DNA measurement using nucleic acid hybridization techniques are known to those of skill in the art (see Sambrook et al., *supra*). Hybridization based assays refer to assays in which a probe nucleic acid is hybridized to a target nucleic acid. Usually the nucleic acid hybridization probes of the invention are entirely or substantially identical to a contiguous sequence of the
20 hTERT gene. Preferably, nucleic acid probes are at least about 10 bases, often at least about 20 bases, and sometimes at least about 200 bases or more in length. Methods of selecting nucleic acid probe sequences for use in nucleic acid hybridization are discussed in Sambrook et al., *supra*. In some formats, at least one of the target and probe is immobilized. The immobilized nucleic acid may be DNA, RNA, or another
25 oligo- or poly-nucleotide, and may comprise natural or non-naturally occurring nucleotides, nucleotide analogs, or backbones. Such assays may be in any of several formats including: Southern, Northern, dot and slot blots, high-density polynucleotide or oligonucleotide arrays (e.g., GeneChips™ Affymetrix), dip sticks, pins, chips, or beads. All of these techniques are well known in the art and are the
30 basis of many commercially available diagnostic kits. Hybridization techniques are generally described in Hames et al., ed., NUCLEIC ACID HYBRIDIZATION, A

PRACTICAL APPROACH IRL Press, (1985); Gall and Pardue *Proc. Natl. Acad. Sci., U.S.A.*, 63: 378-383 (1969); and John et al., *Nature*, 223: 582-587 (1969).

A variety of nucleic acid hybridization formats are known to those skilled in the art. For example, one common format is direct hybridization, in which a target nucleic acid is hybridized to a labeled, complementary probe. Typically, labeled nucleic acids are used for hybridization, with the label providing the detectable signal. A useful method for evaluating the presence, absence, or quantity of DNA encoding hTRT proteins in a sample involves a Southern transfer of DNA from a sample and hybridization of a labeled hTRT specific nucleic acid probe.

Other common hybridization formats include sandwich assays and competition or displacement assays. Sandwich assays are commercially useful hybridization assays for detecting or isolating nucleic acid sequences. Such assays utilize a "capture" nucleic acid covalently immobilized to a solid support and a labeled "signal" nucleic acid in solution. The biological or clinical sample will provide the target nucleic acid. The "capture" nucleic acid and "signal" nucleic acid probe hybridize with the target nucleic acid to form a "sandwich" hybridization complex. To be effective, the signal nucleic acid cannot hybridize with the capture nucleic acid.

b) CHIP-BASED AND SLIDE-BASED ASSAYS

The present invention also provides probe-based hybridization assays employing arrays of immobilized oligonucleotide or polynucleotides to which an hTRT nucleic acid can hybridize (i.e., to some, but usually not all or even most, of the immobilized oligo- or poly-nucleotides). High density oligonucleotide arrays or polynucleotide arrays provide a means for efficiently detecting the presence and characteristics (e.g., sequence) of a target nucleic acid (i.e., hTRT gene).

Techniques are known for producing arrays containing thousands of oligonucleotides complementary to defined sequences, at defined locations on a surface using photolithographic techniques for synthesis *in situ* (see, e.g., U.S. Patent Nos.

5,578,832; 5,556,752; and 5,510,270; Fodor et al., 1991, *Science* 251:767; Pease et al., 1994, *Proc. Natl. Acad. Sci. USA* 91:5022; and Lockhart et al., 1996, *Nature*

Biotech 14:1675) or other methods for rapid synthesis and deposition of defined oligonucleotides (Blanchard et al., 1996, *Biosensors & Bioelectronics* 11:687).

When these methods are used, oligonucleotides (e.g., 20-mers) of known sequence are synthesized directly on a surface such as a derivatized glass slide. Usually, the
5 array produced is redundant, having several oligonucleotide probes on the chip specific for the hTERT polynucleotide to be detected.

Combinations of oligonucleotide probes can be designed to detect alternatively spliced mRNAs, or to identify which of various hTERT alleles is expressed in a particular sample.

10 By way of illustration (but not part of the invention) cDNA prepared by reverse transcription of total RNA from a test cell is amplified (e.g., using PCR). Typically the amplification product is labeled, e.g., by incorporation of a fluorescently labeled dNTP. The labeled cDNAs are then hybridized to a chip comprising oligonucleotide probes complementary to various subsequences of the
15 hTERT gene. The positions of hybridization are determined (e.g., in accordance with the general methods of Shalon et al., 1996, *Genome Research* 6:639 or Schena et al., 1996, *Genome Res.* 6:639), and sequence (or other information) deduced from the hybridization pattern, by means well known in the art.

By way of another illustration (but not part of the invention) two
20 cDNA samples, each labeled with a different fluorescent group, are hybridized to the same chip. The ratio of the hybridization of each labeled sample to sites complementary to the hTERT gene are then assayed. If both samples contain the same amount of hTERT mRNA, the ratio of the two fluors will be 1:1 (it will be appreciated that the signal from the fluors may need to be adjusted to account for any difference
25 in the molar sensitivity of the fluors). In contrast, if one sample is from a healthy (or control) tissue and the second sample is from a cancerous tissue the fluor used in the second sample will predominate.

c) IN SITU HYBRIDIZATION

30 *In situ* hybridization assays are well known and are generally described in Angerer et al., *METHODS ENZYMOL.*, 152: 649-660 (1987) and Ausubel et al.,

supra. In an *in situ* hybridization assay, cells or tissue specimens are fixed to a solid support, typically in a permeabilized state, typically on a glass slide. The cells are then contacted with a hybridization solution at a moderate temperature to permit annealing of labeled nucleic acid probes (e.g., ³⁵S-labeled riboprobes, fluorescently labeled probes) completely or substantially complementary to hTERT. Free probe is removed by washing and/or nuclease digestion, and bound probe is visualized directly on the slide by autoradiography or an appropriate imaging techniques, as is known in the art.

4) SPECIFIC DETECTION OF VARIANTS

Variant hTERT genes that can be detected include those characterized by premature stop codons, deletions, substitutions or insertions. Deletions can be detected by the decreased size of the gene, mRNA transcript, or cDNA. Similarly, insertions can be detected by the increased size of the gene, mRNA transcript, or cDNA. Insertions and deletions could also cause shifts in the reading frame that lead to premature stop codons or longer open reading frames. Substitutions, deletions, and insertions can also be detected by probe hybridization. Alterations can also be detected by observing changes in the size of the variant hTERT polypeptide (e.g., by Western analysis) or by hybridization or specific amplification as appropriate. Alternatively, mutations can be determined by sequencing of the gene or gene product according to standard methods. In addition, and as noted above, amplification assays and hybridization probes can be selected to target particular abnormalities specifically. For example, nucleic acid probes or amplification primers can be selected that specifically hybridize to or amplify, respectively, the region encompassing the deletion, substitution, or insertion. Where the hTERT gene harbors such a mutation, the probe will either (1) fail to hybridize or the amplification reaction will fail to provide specific amplification or cause a change in the size of the amplification product or hybridization signal; or (2) the probe or amplification reaction encompasses the entire deletion or either end of the deletion (deletion junction); or (3) similarly, probes and amplification primers can be selected that specifically target point mutations or insertions.

5) DETECTION OF MUTANT hTERT ALLELES

Mutations in the hTERT gene can be responsible for disease initiation or can contribute to a disease condition. Alterations of the genomic DNA of hTERT can affect levels of gene transcription, change amino acid residues in the hTERT protein, cause truncated hTERT polypeptides to be produced, alter pre-mRNA processing pathways (which can alter hTERT mRNA levels), and cause other consequences as well.

Alterations of genomic DNA in non-hTERT loci can also affect expression of hTERT or telomerase by altering the enzymes or cellular processes that are responsible for regulating hTERT, hTR, and telomerase-associated protein expression and processing and RNP assembly and transport. Alterations which affect hTERT expression, processing, or RNP assembly could be important for cancer progression, for diseases of aging, for DNA damage diseases, and others.

Detection of mutations in hTERT mRNA or its gene and gene control elements can be accomplished in accordance with the methods herein in multiple ways. Illustrative examples include the following: A technique termed primer screening can be employed; PCR primers are designed whose 3' termini anneal to nucleotides in a sample DNA (or RNA) that are possibly mutated. If the DNA (or RNA) is amplified by the primers, then the 3' termini matched the nucleotides in the gene; if the DNA is not amplified, then one or both termini did not match the nucleotides in the gene, indicating a mutation was present. Similar primer design can be used to assay for point mutations using the Ligase Chain Reaction (LCR, described *supra*). Restriction fragment length polymorphism, RFLP (Pourzand, C., Cerutti, P. (1993) *Mutat. Res* 288: 113-121), is another technique that can be applied in the present method. A Southern blot of human genomic DNA digested with various restriction enzymes is probed with an hTERT specific probe. Differences in the fragment number or sizes between the sample and a control indicate an alteration of the experimental sample, usually an insertion or deletion. Single strand conformation polymorphism, SSCP (Orrita, M., et al. (1989) *PNAS USA* 86:2766-70), is another technique that can be applied in the present method. SSCP is

based on the differential migration of denatured wild-type and mutant single-stranded DNA (usually generated by PCR). Single-stranded DNA will take on a three-dimensional conformation that is sequence-specific. Sequence differences as small as a single base change can result in a mobility shift on a nondenaturing gel.

5 SSCP is one of the most widely used mutation screening methods because of its simplicity. Denaturing Gradient Gel Electrophoresis, DGGE (Myers, R. M., Maniatis, T. and Lerman, L., (1987) *Methods in Enzymology*, 155: 501-527), is another technique that can be applied in the present method. DGGE identifies mutations based on the melting behavior of double-stranded DNA. Specialized

10 denaturing electrophoresis equipment is utilized to observe the melting profile of experimental and control DNAs: a DNA containing a mutation will have a different mobility compared to the control in these gel systems. The examples discussed illustrate commonly employed methodology; many other techniques exist which are known by those skilled in the art and can be applied in accordance with the teachings

15 herein.

F) KARYOTYPE ANALYSIS

The present invention further provides methods and reagents for karyotype or other chromosomal analysis using hTERT-sequence probes and/or

20 detecting or locating hTERT gene sequences in chromosomes from a human patient, human cell line, or non-human cell. In one embodiment, amplification (i.e., change in copy number), deletion (i.e., partial deletion), insertion, substitution, or changes in the chromosomal location (e.g., translocation) of an hTERT gene may be correlated with the presence of a pathological condition or a predisposition to developing a

25 pathological condition (e.g., cancer).

It has been determined by the present inventors that, in normal human cells, the hTERT gene maps close to the telomere of chromosome 5p (see Example 5, *infra*). The closest STS marker is D5S678 (see Figure 8). The location can be used to identify markers that are closely linked to the hTERT gene. The markers can be

30 used to identify YACs, STSs, cosmids, BACs, lambda or P1 phage, or other clones which contain hTERT genomic sequences or control elements. The markers or the

gene location can be used to scan human tissue samples for alterations in the normal hTERT gene location, organization or sequence that is associated with the occurrence of a type of cancer or disease. This information can be used in a diagnostic or prognostic manner for the disease or cancer involved. Moreover, the nature of any alterations to the hTERT gene can be informative as to the nature by which cells become immortal. For instance, a translocation event could indicate that activation of hTERT expression occurs in some cases by replacing the hTERT promoter with another promoter which directs hTERT transcription in an inappropriate manner. Methods and reagents of the invention of this type can be used to inhibit hTERT activation. The location may also be useful for determining the nature of hTERT gene repression in normal somatic cells, for instance, whether the location is part of non-expressing heterochromatin. Nuclease hypersensitivity assays for distinguishing heterochromatin and euchromatin are described, for example, in Wu et al., 1979, *Cell* 16:797; Groudine and Weintraub, 1982, *Cell* 30:131 Gross and Garrard, 1988, *Ann. Rev. Biochem.* 57:159.

In one embodiment, alterations to the hTERT gene are identified by karyotype analysis, using any of a variety of methods known in the art. One useful technique is *in situ* hybridization (ISH). Typically, when *in situ* hybridization techniques are used for karyotype analysis, a detectable or detectably-labeled probe is hybridized to a chromosomal sample *in situ* to locate an hTERT gene sequence. Generally, ISH comprises one or more of the following steps: (1) fixation of the tissue, cell or other biological structure to be analyzed; (2) prehybridization treatment of the biological structure to increase accessibility of target DNA (e.g., denaturation with heat or alkali), and to reduce nonspecific binding (e.g., by blocking the hybridization capacity of repetitive sequences, e.g., using human genomic DNA); (3) hybridization of one or more nucleic acid probes (e.g., conventional nucleic acids, PNAs, or probes containing other nucleic acid analogs) to the nucleic acid in the biological structure or tissue; (4) posthybridization washes to remove nucleic acid fragments not bound in the hybridization; and, (5) detection of the hybridized nucleic acid fragments. The reagents used in each of these steps and conditions for their use vary depending on the particular application. It will be appreciated that these steps

can be modified in a variety of ways well known to those of skill in the art.

In one embodiment of ISH, the hTERT probe is labeled with a fluorescent label (fluorescent *in situ* hybridization; "FISH"). Typically, it is desirable to use dual color fluorescent *in situ* hybridization, in which two probes are
5 utilized, each labeled by a different fluorescent dye. A test probe that hybridizes to the hTERT sequence of interest is labeled with one dye, and a control probe that hybridizes to a different region is labeled with a second dye. A nucleic acid that hybridizes to a stable portion of the chromosome of interest, such as the centromere region, can be used as the control probe. In this way, one can account for
10 differences between efficiency of hybridization from sample to sample.

The ISH methods for detecting chromosomal abnormalities (e.g., FISH) can be performed on nanogram quantities of the subject nucleic acids. Paraffin embedded normal tissue or tumor sections can be used, as can fresh or frozen material, tissues, or sections. Because FISH can be applied to limited
15 material, touch preparations prepared from uncultured primary tumors can also be used (*see, e.g.,* Kallioniemi et al., 1992, *Cytogenet. Cell Genet.* 60:190). For instance, small biopsy tissue samples from tumors can be used for touch preparations (*see, e.g.,* Kallioniemi et al., *supra*). Small numbers of cells obtained from aspiration biopsy or cells in bodily fluids (*e.g.,* blood, urine, sputum and the like)
20 can also be analyzed. For prenatal diagnosis, appropriate samples will include amniotic fluid, maternal blood, and the like. Useful hybridization protocols applicable to the methods and reagents disclosed here are described in Pinkel et al., 1988, *Proc. Natl. Acad. Sci. USA*, 85:9138; EPO Pub. No. 430,402; Choo, ed., METHODS IN MOLECULAR BIOLOGY VOL. 33: IN SITU HYBRIDIZATION PROTOCOLS,
25 Humana Press, Totowa, New Jersey, (1994); and Kallioniemi et al., *supra*.

Other techniques useful for karyotype analysis include, for example, techniques such as quantitative Southern blotting, quantitative PCR, or comparative genomic hybridization (Kallioniemi et al., 1992, *Science*, 258:818), using the hTERT probes and primers of the invention which may be used to identify amplification,
30 deletion, insertion, substitution or other rearrangement of hTERT sequences in chromosomes in a biological sample.

Reference is made herein to parent UK Patent Application No 97
208890.4 published under Serial No 2317891 for details of assays for TRT
polypeptides, in particular; electrophoretic assays; immunoassays (competitive and
non-competitive); other assay formats; substrates, solid supports, membranes and
5 filters. Also for details of anti-TRT antibody assays, assay combinations and assay
kits.

G) ASSAY COMBINATIONS

The diagnostic and prognostic assays described herein can be carried
10 out in various combinations and can also be carried out in conjunction with other
diagnostic or prognostic tests. For example, when the present methods are used to
detect the presence of cancer cells in patient sample, the presence of hTRT can be
used to determine the stage of the disease, whether a particular tumor is likely to
invade adjoining tissue or metastasize to a distant location, and whether a recurrence
15 of the cancer is likely. Tests that may provide additional information include
microscopic analysis of biopsy samples, detection of antigens (e.g., cell-surface
markers) associated with tumorigenicity (e.g., using histocytochemistry, FACS, or
the like), imaging methods (e.g., upon administration to a patient of labeled anti-
tumor antibodies), telomerase activity assays, telomere length assays, hTR assays, or
20 the like. Such combination tests can provide useful information regarding the
progression of a disease.

H) KITS

The present invention also provides kits useful for the screening,
25 monitoring, diagnosis and prognosis of patients with a telomerase-related condition,
or for determination of the level of expression of hTRT in cells or cell lines. The
kits include one or more reagents for quantifying expression of the hTRT gene.
Preferred reagents include nucleic acid primers and probes that specifically bind to
the hTRT gene, or portions thereof, along with proteins, peptides, antibodies, and
30 control primers, probes, oligonucleotides, proteins, peptides and antibodies. Other
materials, including enzymes, buffers, reagents (labels, dNTPs), may be included.

The kits may include alternatively, or in combination with any of the other components described herein, an antibody that specifically binds to hTERT polypeptides or subsequences thereof. The antibody can be monoclonal or polyclonal. The antibody can be conjugated to another moiety such as a label and/or it can be immobilized on a solid support (substrate). The kit(s) may also contain a second antibody for detection of hTERT polypeptide/antibody complexes or for detection of hybridized nucleic acid probes, as well as one or more hTERT peptides or proteins for use as control or other reagents.

The antibody or hybridization probe may be free or immobilized on a solid support such as a test tube, a microtiter plate, a dipstick and the like. The kit may also contain instructional materials teaching the use of the antibody or hybridization probe in an assay for the detection of TRT. The kit may contain appropriate reagents for detection of labels, or for labeling positive and negative controls, washing solutions, dilution buffers and the like.

15

IX) IDENTIFICATION OF MODULATORS OF TELOMERASE ACTIVITY

A) GENERALLY

The invention provides compounds and treatments that modulate the activity or expression of a telomerase or telomerase component (e.g., hTERT protein). The invention also provides assays and screening methods (including high-throughput screens) for identification of compounds and treatments that modulate telomerase activity or expression. These modulators of telomerase activity and expression (hereinafter referred to as "modulators") include telomerase agonists (which increase telomerase activity and/or expression) and telomerase antagonists (which decrease telomerase activity and/or expression).

The modulators of the invention have a wide variety of uses. For example, it is contemplated that telomerase modulators will be effective therapeutic agents for treatment of human diseases. Screening for agonist activity and transcriptional activators provides for compositions that increase telomerase activity in a cell (including a telomere dependent replicative capacity, or a "partial" telomerase

activity). Such agonist compositions provide for methods of immortalizing otherwise normal untransformed cells, including cells which can express useful proteins. Such agonists can also provide for methods of controlling cellular senescence. Conversely, screening for antagonist activity provides for compositions that decrease telomere dependent replicative capacity, thereby mortalizing otherwise immortal cells, such as cancer cells. Screening for antagonist activity provides for compositions that decrease telomerase activity, thereby preventing unlimited cell division of cells exhibiting unregulated cell growth, such as cancer cells. Illustrative diseases and conditions that may be treated using modulators are listed herein, e.g., in Sections VI and VIII, *supra*.

10 In general, the modulators of the invention can be used whenever it is desired to increase or decrease a telomerase activity in a cell or organism. Thus, in addition to use in treatment of disease, a modulator that increases hTERT expression levels can be used to produce a cultured human cell line having properties as generally described in Section VII, *supra*, and various other uses that will be apparent to one of skill.

15 A compound or treatment modulates "expression" of telomerase or a telomerase component when administration of the compound or treatment changes the rate or level of transcription of the gene encoding a telomerase component (e.g., the gene encoding hTERT mRNA). A compound or treatment affects a telomerase "activity" when administration of the compound or treatment changes a telomerase activity such as

20 any activity described in parent UK Patent Application No 97 208890.4 published under Serial No 2317891 (e.g., including processive or non-processive telomerase catalytic activity; telomerase processivity; conventional reverse transcriptase activity; nucleolytic activity; primer or substrate binding activity; dNTP binding activity; RNA binding activity; telomerase RNP assembly; and protein binding activity). It

25 will be appreciated that there is not necessarily a sharp delineation between changes in "activity" and changes in "expression," and that these terms are used for ease of discussion and not for limitation. It will also be appreciated that the modulators of the invention should specifically affect telomerase activity or expression (e.g., without generally changing the expression of housekeeping proteins such as actin) rather than,

30 for example, reducing expression of a telomerase component by nonspecific poisoning of a target cell.

B) ASSAYS FOR IDENTIFICATION OF TELOMERASE MODULATORS

The invention provides methods and reagents to screen for
5 compositions or compounds capable of affecting expression of a telomerase or telomerase component. Thus, the present invention provides assays that can be used to screen for agents that increase the activity of telomerase, for example, by causing hTERT protein or telomerase to be expressed in a cell in which it normally is not expressed or by increasing telomerase activity levels in telomerase positive cells.

10 For details of hTERT activities and assays therefor, reference is made to parent UK Patent Application No 97 208890.4 published under Serial No 2317891. The invention contemplates screening for compositions that increase or decrease the transcription of the hTERT gene.

In one embodiment, an assay for identification of modulators comprises
15 contacting one or more cells (i.e., "test cells") with a test compound, and determining whether the test compound affects expression or activity of a telomerase (or telomerase component) in the cell. Usually this determination comprises comparing the activity or expression in the test cell compared to a similar cell or cells (i.e., control cells) that have not been contacted with the test compound. Alternatively, cell extracts may be
20 used in place of intact cells. In a related embodiment, the test compound is administered to a multicellular organism (e.g., a plant or animal). The telomerase or telomerase component may be wholly endogenous to the cell or multicellular organism (i.e., encoded by naturally occurring endogenous genes), or may be a recombinant cell or transgenic organism comprising one or more recombinantly expressed telomerase
25 components (e.g., hTERT, hTR, telomerase-associated proteins), or may have both endogenous and recombinant components. Thus, in one embodiment, telomerase-activity-modulators are administered to mortal cells. In another embodiment, telomerase-activity-modulators are administered to immortal cells. For example, antagonists of telomerase-mediated DNA replication can be identified by
30 administering the putative inhibitory composition to a cell that is known to exhibit significant amounts of telomerase activity, such as cancer cells, and measuring

whether a decrease in telomerase activity, telomere length, or proliferative capacity is observed, all of which are indicative of a compound with antagonist activity.

In another embodiment, the modulator is identified by monitoring a change in expression of a TRT gene product (e.g., RNA or protein) in a cell, animal,
5 *in vitro* expression system, or other expression system.

In another embodiment, the modulator is identified by changing the expression of a reporter gene, such as that described in Example 15, whose expression is regulated, in whole or part, by a naturally occurring TRT regulatory element such as a promoter or enhancer. In a related embodiment, the ability of a
10 test compound to bind to a telomerase component (e.g., hTERT), RNA, or gene regulatory sequence (e.g., the TRT gene promoter) is assayed.

Assay formats for identification of compounds that affect expression and activity of proteins are well known in the biotechnological and pharmaceutical industries, and numerous additional assays and variations of the illustrative assays
15 provided *supra* will be apparent to those of skill.

Changes in telomerase activity or expression can be measured by any suitable method. Changes in levels of expression of a telomerase component (e.g., hTERT protein) or precursor (e.g., hTERT mRNA) can be assayed using methods well known to those of skill, some of which are described hereinabove, e.g., in Section
20 VIII and including monitoring levels of TRT gene products (e.g., protein and RNAs) by hybridization (e.g., using TRT probes and primers), immunoassays (e.g., using anti-TRT antibodies), RNase protection assays, amplification assays, or any other suitable detection means described herein or known in the art. Quantitating amounts of nucleic acid in a sample (e.g., evaluating levels of RNA, e.g., hTR or hTERT
25 mRNA) is also useful in evaluating *cis*- or *trans*- transcriptional regulators.

Similarly, changes in telomerase activity can be measured using methods such as those described in parent UK Patent Application No 97 208890.4 published under Serial No 2317891 or other assays of telomerase function. Quantitation of telomerase activity, when desired, may be carried out by any method.
30 Telomerase antagonists that can cause or accelerate loss of telomeric structure can be identified by monitoring and measuring their effect on telomerase activity *in vivo*, *ex*

vivo, or *in vitro*, or by their effects on telomere length (as measured or detected through staining, use of tagged hybridization probes or other means) or, simply, by the inhibition of cell division of telomerase positive cancer cells (critical shortening of telomeres leads to a phenomenon termed "crisis" or M2 senescence (Shay, 1991) *Biochem. Biophys. Acta* 1072:1-7), which cancer cells have bypassed by the activation of telomerase, but which, in the absence of telomerase, will lead to their senescence or death through chromosomal deletion and rearrangement). The *in vivo* human telomerase activity reconstitution provides for a method of screening for telomerase modulators in cells or animals from any origin. Such agonists can be identified in an activity assay of the invention, including measurements of changes in telomere length. Other examples of assays measuring telomerase activity in cells include assays for the accumulation or loss of telomere structure, the TRAP assay or a quantitative polymerase chain reaction assay.

In one embodiment, the assays of the invention also include a method where the test compound produces a statistically significant decrease in the activity of hTERT as measured by the incorporation of a labeled nucleotide into a substrate compared to the relative amount of incorporated label in a parallel reaction lacking the test compound, thereby determining that the test compound is a telomerase inhibitor.

The methods of the invention are amenable to adaptations from protocols described in the scientific and patent literature and known in the art. For example, when a telomerase or TERT protein of this invention is used to identify compositions which act as modulators of telomerase activities, large numbers of potentially useful molecules can be screened in a single test. The modulators can have an inhibitory (antagonist) or potentiating (agonist) effect on telomerase activity. For example, if a panel of 1,000 inhibitors is to be screened, all 1,000 inhibitors can potentially be placed into one microtiter well and tested simultaneously. If such an inhibitor is discovered, then the pool of 1,000 can be subdivided into 10 pools of 100 and the process repeated until an individual inhibitor is identified.

In drug screening large numbers of compounds are examined for their ability to act as telomerase modulators, a process greatly accelerated by the

techniques of high throughput screening. The assays for telomerase activity, full or partial, described herein may be adapted to be used in a high throughput technique. Those skilled in the art appreciate that there are numerous methods for accomplishing this purpose.

5 Additional methods for identifying modulators of a telomerase activity have been described in U.S. Patent No. 5,645,986, which is incorporated herein by reference. Reagents such as hTRT polynucleotides, probes and primers, highly purified hTR, hTRT and telomerase, as well as anti-telomerase and anti-TRT antibodies, may all be used in assays, e.g., as controls, standards, binding or
10 hybridization agents, or otherwise.

It will be recognized that recombinantly produced telomerase and TRT (e.g., hTRT) will be useful in assays for identification of modulators. Such a screening assay can utilize telomerase or hTRT derived by a full or partial reconstitution of telomerase activity, or by an augmentation of existing activity. Such
15 an assay or screens can be used to test for the ability of telomerase to synthesize telomeric DNA or to test for any one or all or of the "partial activities" of hTRT and TRTs generally. The assay can incorporate *ex vivo* modification of cells which have been manipulated to express telomerase with or without its RNA moiety or associated proteins, and these can be re-implanted into an animal, which can be used for *in vivo*
20 testing. *In vivo* assays systems can employ "knockout" cells, in which one or several units of the endogenous telomerase enzyme complex have been deleted or inhibited, as well as cells in which an exogenous or endogenous telomerase activity is reconstituted or activated.

25 C) EXEMPLARY TELOMERASE MODULATORS

1) GENERALLY

The test compounds referred to *supra* may be any of a large variety of compounds, both naturally occurring and synthetic, organic and inorganic, and including polymers (e.g., oligopeptides, polypeptides, oligonucleotides, and
30 polynucleotides), small molecules, antibodies (as broadly defined herein), sugars, fatty acids, nucleotides and nucleotide analogs, analogs of naturally occurring

structures (e.g., peptide mimetics, nucleic acid analogs, and the like), and numerous other compounds.

The invention provides modulators of all types, without limitation to any particular mechanism of action. For illustrative purposes, examples of modulators
5 include compounds or treatments that increase or decrease expression of a telomerase component gene product (e.g., products of the hTERT gene), including change the rate or level of transcription of the TRT gene, or translation, transport or stability of a gene product, or the like, by binding to the gene or gene product (e.g., by interacting with a factor (e.g., a transcription regulatory protein) that affects transcription of the hTERT
10 gene or another telomerase component).

2) PEPTIDE MODULATORS

Potential modulators of telomerase activity also include peptides (e.g., inhibitory (antagonist) and activator (agonist) peptide modulators). Reference is made
15 to parent UK Patent Application No 97208890.4 published under Serial No 2317891 for further details.

3) INHIBITORY NATURAL COMPOUNDS AS MODULATORS OF TELOMERASE ACTIVITY

20 In addition, a large number of potentially useful activity-modifying compounds can be screened in extracts from natural products as a source material. Sources of such extracts can be from a large number of species of fungi, actinomyces, algae, insects, protozoa, plants, and bacteria. Those extracts showing inhibitory activity can then be analyzed to isolate the active molecule. See for
25 example, Turner (1996) *J. Ethnopharmacol* 51(1-3):39-43; Suh (1995) *Anticancer Res.* 15:233-239.

4) INHIBITORY OLIGONUCLEOTIDES

One particularly useful set of inhibitors provided by the present
30 invention includes oligonucleotides which are able to bind to the hTERT gene, in either case preventing or inhibiting the production of functional hTERT protein.

Oligonucleotides can be generated using the techniques of SELEX (Tuerk, 1997, *Methods Mol Biol* 67, 2190). In this technique a very large pool (106-109) of random sequence nucleic acids is bound to the target using conditions that cause a large amount of discrimination between molecules with high affinity and low affinity for binding the target. The bound molecules are separated from unbound, and the bound molecules are amplified by virtue of a specific nucleic acid sequence included at their termini and suitable amplification reagents. This process is reiterated several times until a relatively small number of molecules remain that possess high binding affinity for the target. These molecules can then be tested for their ability to modulate telomerase activity as described herein.

The inhibitory oligonucleotides of the invention can be transferred into the cell using a variety of techniques well known in the art. For example, oligonucleotides can be delivered into the cytoplasm without specific modification. Alternatively, they can be delivered by the use of liposomes which fuse with the cellular membrane or are endocytosed, *i.e.*, by employing ligands attached to the liposome or directly to the oligonucleotide, that bind to surface membrane protein receptors of the cell resulting in endocytosis. Alternatively, the cells may be permeabilized to enhance transport of the oligonucleotides into the cell, without injuring the host cells. One can use a DNA binding protein, *e.g.*, HBGF-1, known to transport an oligonucleotide into a cell.

D) MODULATOR SYNTHESIS

It is contemplated that the telomerase modulators of the invention will be made using methods well known in the pharmaceutical arts, including combinatorial methods and rational drug design techniques.

1) COMBINATORIAL CHEMISTRY METHODOLOGY

The creation and simultaneous screening of large libraries of synthetic molecules can be carried out using well-known techniques in combinatorial chemistry, for example, see van Breemen (1997) *Anal Chem* 69:2159-2164; Lam (1997) *Anticancer Drug Des* 12:145-167 (1997).

As noted above, combinatorial chemistry methodology can be used to create vast numbers of oligonucleotides (or other compounds) that can be rapidly screened for specific oligonucleotides (or compounds) that have appropriate binding affinities and specificities toward any target can be utilized (for general background information Gold (1995) *J. of Biol. Chem.* 270:13581-13584).

2) RATIONAL DRUG DESIGN

Rational drug design involves an integrated set of methodologies that include structural analysis of target molecules, synthetic chemistries, and advanced computational tools. When used to design modulators, the objective of rational drug design is to understand a molecule's three-dimensional shape and chemistry. Rational drug design is aided by X-ray crystallographic data or NMR data, which can now be determined. Calculations on electrostatics, hydrophobicities and solvent accessibility is also helpful. See, for example, Coldren (1997) *Proc. Natl. Acad. Sci. USA* 94:6635-6640.

E) KITS

The invention also provides kits that can be used to aid in determining whether a test compound is a modulator of a TRT activity. The kit will typically include one or more of the following components: a substantially purified TRT polynucleotide (including probes and primers); a plasmid capable of expressing a TRT (e.g., hTRT) when introduced into a cell or cell-free expression system; a plasmid capable of expressing a TR (e.g., hTR) when introduced into a cell or cell-free expression system; cells or cell lines; a composition to detect a change in TRT activity; and, an instructional material teaching a means to detect and measure a change in the TRT activity, indicating that a change in the telomerase activity in the presence of the test compound is an indicator that the test compound modulates the telomerase activity, and one or more containers. The kit can also include means, such as TRAP assay reagents or reagents for a quantitative polymerase chain reaction assay, to measure a change in TRT activity. The kit may also include instructional material teaching a means to detect and measure a change in the TRT activity, indicating

that a change in the telomerase activity in the presence of the test compound is an indicator that the test compound modulates the telomerase activity.

X) TRANSGENIC ORGANISMS (TELOMERASE KNOCKOUT CELLS AND ANIMAL MODELS)

5 The invention also provides transgenic non-human multicellular organisms (e.g., plants and non-human animals) or unicellular organisms (e.g., yeast) comprising an exogenous TRT gene sequence, which may be a coding sequence or a regulatory (e.g., promoter) sequence. In one embodiment, the organism expresses an
10 exogenous TRT polypeptide, having a sequence of a human TRT protein. In a related embodiment, the organism also expresses a telomerase RNA component (e.g., hTR).

The invention also provides unicellular and multicellular organisms (or cells therefrom) in which at least one gene encoding a telomerase component (e.g., TRT or TR) or telomerase-associated protein is mutated or deleted
15 (i.e., in a coding or regulatory region) such that native telomerase is not expressed, or is expressed at reduced levels or with different activities when compared to wild-type cells or organisms. Such cells and organisms are often referred to as "gene knock-out" cells or organisms.

The invention further provides cells and organisms in which an
20 endogenous telomerase gene (e.g., murine TRT) is either present or optionally mutated or deleted and an exogenous telomerase gene or variant (e.g., human TRT) is introduced and expressed. Cells and organisms of this type will be useful, for example, as model systems for identifying modulators of hTRT activity or expression; determining the effects of mutations in telomerase component genes, and
25 other uses such as determining the developmental timing and tissue location of telomerase activity (e.g., for assessing when to administer a telomerase modulator and for assessing any potential side effects).

Examples of multicellular organisms include plants, insects, and nonhuman animals such as mice, rats, rabbits, monkeys, apes, pigs, and other
30 nonhuman mammals. An example of a unicellular organism is a yeast.

Methods for alteration or disruption of specific genes (e.g., endogenous

TRT genes) are well known to those of skill, see, e.g., Baudin et al., 1993, *Nucl. Acids Res.* 21:3329; Wach et al., 1994, *Yeast* 10:1793; Rothstein, 1991, *Methods Enzymol.* 194:281; Anderson, 1995, *Methods Cell Biol.* 48:31; Pettitt et al., 1996, *Development* 122:4149-4157; Ramirez-Solis et al., 1993, *Methods Enzymol.* 225:855; 5 and Thomas et al., 1987, *Cell* 51:503, each of which is incorporated herein by reference in its entirety for all purposes.

The "knockout" cells and animals of the invention include cells and animals in which one or several units of the endogenous telomerase enzyme complex have been deleted or inhibited. Reconstitution of telomerase activity will save the 10 cell or animal from senescence or, for cancer cells, cell death caused by its inability to maintain telomeres. Methods of altering the expression of endogenous genes are well known to those of skill in the art. Typically, such methods involve altering or replacing all or a portion of the regulatory sequences controlling expression of the particular gene to be regulated. The regulatory sequences, e.g., the native promoter 15 can be altered. The conventional technique for targeted mutation of genes involves placing a genomic DNA fragment containing the gene of interest into a vector, followed by cloning of the two genomic arms associated with the targeted gene around a selectable neomycin-resistance cassette in a vector containing thymidine kinase. This "knock-out" construct is then transfected into the appropriate host cell, 20 i.e., a mouse embryonic stem (ES) cell, which is subsequently subjected to positive selection (using G418, for example, to select for neomycin-resistance) and negative selection (using, for example, FIAU to exclude cells lacking thymidine kinase), allowing the selection of cells which have undergone homologous recombination with the knockout vector. This approach leads to inactivation of the gene of interest. See, 25 e.g., U.S. patents 5,464,764; 5,631,153; 5,487,992; and, 5,627,059.

"Knocking out" expression of an endogenous gene can also be accomplished by the use of homologous recombination to introduce a heterologous nucleic acid into the regulatory sequences (e.g., promoter) of the gene of interest. To prevent expression of functional enzyme or product, simple mutations that either alter 30 the reading frame or disrupt the promoter can be suitable. To up-regulate expression, a native promoter can be substituted with a heterologous promoter that induces higher

levels of transcription. Also, "gene trap insertion" can be used to disrupt a host gene, and mouse ES cells can be used to produce knockout transgenic animals, as described for example, in Holzschu (1997) *Transgenic Res* 6: 97-106.

- Altering the expression of endogenous genes by homologous recombination can also be accomplished by using nucleic acid sequences comprising the structural gene in question. Upstream sequences are utilized for targeting heterologous recombination constructs. Utilizing TRT structural gene sequence information, such as SEQUENCE ID NO:1, one of skill in the art can create homologous recombination constructs with only routine experimentation.
- 10 Homologous recombination to alter expression of endogenous genes is described in U.S. Patent 5,272,071, and WO 91/09955, WO 93/09222, WO 96/29411, WO 95/31560, and WO 91/12650. Homologous recombination in mycobacteria is described by Azad (1996) *Proc. Natl. Acad. Sci. USA* 93:4787; Baulard (1996) *J. Bacteriol.* 178:3091; and Pelicic (1996) *Mol. Microbiol.* 20:919. Homologous
- 15 recombination in animals has been described by Moynahan (1996) *Hum. Mol. Genet.* 5:875, and in plants by Offringa (1990) *EMBO J.* 9:3077.

XI) GLOSSARY

- The following terms are defined *infra* to provide additional guidance to
- 20 one of skill in the practice of the invention: adjuvant, allele (& allelic sequence), amino acids (including hydrophobic, polar, charged), conservative substitution, control elements (& regulatory sequences), derivatized, detectable label, elevated level, epitope, favorable and unfavorable prognosis, fusion protein, gene product, hTR, immortal, immunogen and immunogenic, isolated, modulator, motif, nucleic
- 25 acid (& polynucleotide), oligonucleotides (& oligomers), operably linked, polypeptide, probe (including nucleic acid probes & antibody probes), recombinant, selection system, sequence, specific binding, stringent hybridization conditions (& stringency), substantial identity (& substantial similarity), substantially pure (& substantially purified), telomerase-negative and telomerase-positive cells, telomerase
- 30 catalytic activity, telomerase-related, and test compound.

As used herein, the term "adjuvant" refers to its ordinary meaning of

any substance that enhances the immune response to an antigen with which it is mixed. Adjuvants useful in the present invention include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole
5 limpet hemocyanin, and dinitrophenol. BCG (*Bacillus Calmette-Guerin*) and *Corynebacterium parvum* are potentially useful adjuvants.

As used herein, the terms "allele" or "allelic sequence" refer to an alternative form of a nucleic acid sequence (i.e., a nucleic acid encoding hTRT protein). Alleles result from mutations (i.e., changes in the nucleic acid sequence),
10 and generally produce altered and/or differently regulated mRNAs or polypeptides whose structure and/or function may or may not be altered. Common mutational changes that give rise to alleles are generally ascribed to natural deletions, additions, or substitutions of nucleotides that may or may not affect the encoded amino acids. Each of these types of changes may occur alone, in combination with the others, or
15 one or more times within a given gene, chromosome or other cellular nucleic acid. Any given gene may have no, one or many allelic forms. As used herein, the term "allele" refers to either or both a gene or an mRNA transcribed from the gene.

As used herein, "amino acids" are sometimes specified using the standard one letter code: Alanine (A), Serine (S), Threonine (T), Aspartic acid (D),
20 Glutamic acid (E) Asparagine (N), Glutamine (Q), Arginine (R), Lysine (K), Isoleucine (I), Leucine (L), Methionine (M), Valine (V), Phenylalanine (F), Tyrosine (Y), Tryptophan (W), Proline (P), Glycine (G), Histidine (H), Cysteine (C). Synthetic and non-naturally occurring amino acid analogues (and/or peptide linkages) are included.

25 As used herein, "Hydrophobic amino acids" refers to A, L, I, V, P, F, W, and M. As used herein, "polar amino acids" refers to G, S, T, Y, C, N, and Q. As used herein, "charged amino acids" refers to D, E, H, K, and R.

As used herein, "conservative substitution", when describing a protein refers to a change in the amino acid composition of the protein that does not
30 substantially alter the protein's activity. Thus, "conservatively modified variations" of a particular amino acid sequence refers to amino acid substitutions of those amino

acids that are not critical for protein activity or substitution of amino acids with other amino acids having similar properties (e.g., acidic, basic, positively or negatively charged, polar or non-polar, etc.) such that the substitutions of even critical amino acids does not substantially alter activity. Conservative substitution tables providing

5 functionally similar amino acids are well known in the art. The following six groups each contain amino acids that are conservative substitutions for one another: 1) Alanine (A), Serine (S), Threonine (T); 2) Aspartic acid (D), Glutamic acid (E); 3) Asparagine (N), Glutamine (Q); 4) Arginine (R), Lysine (K); 5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); and 6) Phenylalanine (F), Tyrosine (Y),

10 Tryptophan (W) (see also, Creighton (1984) *Proteins*, W.H. Freeman and Company). One of skill in the art will appreciate that the above-identified substitutions are not the only possible conservative substitutions. For example, one may regard all charged amino acids as conservative substitutions for each other whether they are positive or negative. In addition, individual substitutions, deletions

15 or additions which alter, add or delete a single amino acid or a small percentage of amino acids in an encoded sequence can also be "conservatively modified variations". One can also make a "conservative substitution" in a recombinant protein by utilizing one or more codons that differ from the codons employed by the native or wild-type gene. In this instance, a conservative substitution also includes

20 substituting a codon for an amino acid with a different codon for the same amino acid.

As used herein, "control elements" or "regulatory sequences" include enhancers, promoters, transcription terminators, origins of replication, chromosomal integration sequences, 5' and 3' untranslated regions, with which

25 proteins or other biomolecules interact to carry out transcription and translation. For eukaryotic cells, the control sequences will include a promoter and preferably an enhancer, e.g., derived from immunoglobulin genes, SV40, cytomegalovirus, and a polyadenylation sequence, and may include splice donor and acceptor sequences. Depending on the vector system and host utilized, any number of suitable

30 transcription and translation elements, including constitutive and inducible promoters, may be used.

As used herein, a "derivatized" polynucleotide, oligonucleotide, or nucleic acid refers to oligo- and polynucleotides that comprise a derivatized substituent. In some embodiments, the substituent is substantially non-interfering with respect to hybridization to complementary polynucleotides. Derivatized oligo- or polynucleotides that have been modified with appended chemical substituents (e.g., by modification of an already synthesized oligo- or poly-nucleotide, or by incorporation of a modified base or backbone analog during synthesis) may be introduced into a metabolically active eukaryotic cell to hybridize with an hTERT DNA, RNA, or protein where they produce an alteration or chemical modification to a local DNA, RNA, or protein. Alternatively, the derivatized oligo or polynucleotides may interact with and alter hTERT polypeptides, telomerase-associated proteins, or other factors that interact with hTERT DNA or hTERT gene products, or alter or modulate expression or function of hTERT DNA, RNA or protein. Illustrative attached chemical substituents include: europium (III) texaphyrin, cross-linking agents, psoralen, metal chelates (e.g., iron/EDTA chelate for iron catalyzed cleavage), topoisomerases, endonucleases, exonucleases, ligases, phosphodiesterases, photodynamic porphyrins, chemotherapeutic drugs (e.g., adriamycin, doxorubicin), intercalating agents, base-modification agents, immunoglobulin chains, and oligonucleotides. Iron/EDTA chelates are chemical substituents often used where local cleavage of a polynucleotide sequence is desired (Hertzberg et al., 1982, *J. Am. Chem. Soc.* 104: 313; Hertzberg and Dervan, 1984, *Biochemistry* 23: 3934; Taylor et al., 1984, *Tetrahedron* 40: 457; Dervan, 1986, *Science* 232: 464. Illustrative attachment chemistries include: direct linkage, e.g., via an appended reactive amino group (Corey and Schultz (1988) *Science* 238: 1401, which is incorporated herein by reference) and other direct linkage chemistries, although streptavidin/biotin and digoxigenin/anti-digoxigenin antibody linkage methods can also be used. Methods for linking chemical substituents are provided in U.S. Patents 5,135,720, 5,093,245, and 5,055,556, which are incorporated herein by reference. Other linkage chemistries may be used at the discretion of the practitioner.

As used herein, a "detectable label" has the ordinary meaning in the art and refers to an atom (e.g., radionuclide), molecule (e.g., fluorescein), or

complex, that is or can be used to detect (e.g., due to a physical or chemical property), indicate the presence of a molecule or to enable binding of another molecule to which it is covalently bound or otherwise associated. The term "label" also refers to covalently bound or otherwise associated molecules (e.g., a biomolecule such as an enzyme) that act on a substrate to produce a detectable atom, molecule or complex. Detectable labels suitable for use in the present invention include any composition detectable by spectroscopic, photochemical, biochemical, immunochemical, electrical, optical or chemical means. Labels useful in the present invention include biotin for staining with labeled streptavidin conjugate, magnetic beads (e.g., Dynabeads™), fluorescent dyes (e.g., fluorescein, Texas red, rhodamine, green fluorescent protein, enhanced green fluorescent protein, lissamine, phycoerythrin, Cy2, Cy3, Cy3.5, Cy5, Cy5.5, Cy7, FluorX [Amersham], SyBR Green I & II [Molecular Probes], and the like), radiolabels (e.g., ^3H , ^{125}I , ^{35}S , ^{14}C , or ^{32}P), enzymes (e.g., hydrolases, particularly phosphatases such as alkaline phosphatase, esterases and glycosidases, or oxidoreductases, particularly peroxidases such as horse radish peroxidase, and others commonly used in ELISAs), substrates, cofactors, inhibitors, chemiluminescent groups, chromogenic agents, and colorimetric labels such as colloidal gold or colored glass or plastic (e.g., polystyrene, polypropylene, latex, *etc.*) beads. Patents teaching the use of such labels include U.S. Patent Nos. 3,817,837; 3,850,752; 3,939,350; 3,996,345; 4,277,437; 4,275,149; and 4,366,241. Means of detecting such labels are well known to those of skill in the art. Thus, for example, radiolabels and chemiluminescent labels may be detected using photographic film or scintillation counters, fluorescent markers may be detected using a photodetector to detect emitted light (e.g., as in fluorescence-activated cell sorting). Enzymatic labels are typically detected by providing the enzyme with a substrate and detecting the reaction product produced by the action of the enzyme on the substrate, and colorimetric labels are detected by simply visualizing the colored label. Thus, a label is any composition detectable by spectroscopic, photochemical, biochemical, immunochemical, electrical, optical or chemical means. The label may be coupled directly or indirectly to the desired component of the assay according to methods well known in the art. Non-radioactive

labels are often attached by indirect means. Generally, a ligand molecule (e.g., biotin) is covalently bound to the molecule. The ligand then binds to an anti-ligand (e.g., streptavidin) molecule which is either inherently detectable or covalently bound to a signal generating system, such as a detectable enzyme, a fluorescent compound, or a chemiluminescent compound. A number of ligands and anti-ligands can be used. Where a ligand has a natural anti-ligand, for example, biotin, thyroxine, and cortisol, it can be used in conjunction with the labeled, naturally occurring anti-ligands. Alternatively, any haptenic or antigenic compound can be used in combination with an antibody. The molecules can also be conjugated directly to signal generating compounds, e.g., by conjugation with an enzyme or fluorophore. Means of detecting labels are well known to those of skill in the art. Thus, for example, where the label is a radioactive label, means for detection include a scintillation counter, photographic film as in autoradiography, or storage phosphor imaging. Where the label is a fluorescent label, it may be detected by exciting the fluorochrome with the appropriate wavelength of light and detecting the resulting fluorescence. The fluorescence may be detected visually, by means of photographic film, by the use of electronic detectors such as charge coupled devices (CCDs) or photomultipliers and the like. Similarly, enzymatic labels may be detected by providing the appropriate substrates for the enzyme and detecting the resulting reaction product. Also, simple colorimetric labels may be detected by observing the color associated with the label. It will be appreciated that when pairs of fluorophores are used in an assay, it is often preferred that they have distinct emission patterns (wavelengths) so that they can be easily distinguished.

The phrase "elevated level" refers to an amount of hTERT gene product (or other specified substance or activity) in a cell that is elevated or higher than the level in a reference standard, e.g., for diagnosis, the level in normal, telomerase-negative cells in an individual or in other individuals not suffering from the condition, and for prognosis, the level in tumor cells from a variety of grades or classes of, e.g., tumors. As used herein, the term "epitope" has its ordinary meaning of a site on an antigen recognized by an antibody. Epitopes are typically segments of amino acids which are

a small portion of the whole protein. Epitopes may be conformational (*i.e.*, discontinuous). That is, they may be formed from amino acids encoded by noncontiguous parts of a primary sequence that have been juxtaposed by protein folding.

5 The terms "**favorable prognosis**" and "**unfavorable prognosis**" are known in the art. In the context of cancers, "favorable prognosis" means that there is a likelihood of tumor regression or longer survival times for patients with a favorable prognosis relative to those with unfavorable prognosis, whereas "unfavorable prognosis" means that the tumor is likely to be more aggressive, *i.e.*,
10 grow faster and/or metastasize, resulting in a poor outcome or a more rapid course of disease progression for the patient.

 As used herein, the term "**fusion protein**," refers to a composite protein, *i.e.*, a single contiguous amino acid sequence, made up of two (or more) distinct, heterologous polypeptides which are not normally fused together in a single
15 amino acid sequence. Thus, a fusion protein may include a single amino acid sequence that contains two entirely distinct amino acid sequences or two similar or identical polypeptide sequences, provided that these sequences are not normally found together in the same configuration in a single amino acid sequence found in nature. Fusion proteins may generally be prepared using either recombinant nucleic acid
20 methods, *i.e.*, as a result of transcription and translation of a recombinant gene fusion product, which fusion comprises a segment encoding a polypeptide of the invention and a segment encoding a heterologous protein, or by chemical synthesis methods well known in the art. The non-hTRT region(s) of the fusion protein can be fused to the amino terminus of the hTRT polypeptide or the carboxyl terminus, or both or the
25 non-hTRT region can be inserted into the interior of the protein sequence (by moiety inserting or by replacing amino acids) or combinations of the foregoing can be performed.

 As used herein, the term "**gene product**" refers to an RNA molecule transcribed from a gene, or a protein encoded by the gene or translated from the
30 RNA.

 As used herein, "**hTR**" (human telomerase RNA) refers to the RNA

component of human telomerase and any naturally occurring alleles and variants or recombinant variants. hTR is described in detail in U.S. Patent No. 5,583,016 which is incorporated herein by reference in its entirety and for all purposes.

As used herein, the term "immortal," when referring to a cell, has its
5 normal meaning in the telomerase art and refers to cells that have apparently unlimited replicative potential. Immortal can also refer to cells with increased proliferative capacity relative to their unmodified counterparts. Examples of immortal human cells are malignant tumor cells, germ line cells, and certain transformed human cell lines cultured *in vitro* (e.g., cells that have become immortal
10 following transformation by viral oncogenes or otherwise). In contrast, most normal human somatic cells are mortal, i.e., have limited replicative potential and become senescent after a finite number of cell divisions.

As used herein, the terms "immunogen" and "immunogenic" have their ordinary meaning in the art, i.e., an immunogen is a molecule, such as a protein
15 or other antigen, that can elicit an adaptive immune response upon injection into a person or an animal.

As used herein, "isolated," when referring to a molecule or composition, such as, for example, an RNP (e.g., at least one protein and at least one RNA), means that the molecule or composition is separated from at least one other
20 compound, such as a protein, other RNAs, or other contaminants with which it is associated *in vivo* or in its naturally occurring state. Thus, an RNP is considered isolated when the RNP has been isolated from any other component with which it is naturally associated, e.g., cell membrane, as in a cell extract. An isolated composition can, however, also be substantially pure.

25 As used herein, "modulator" refers to any synthetic or natural compound or composition that can change in any way either or both the "full" or any "partial activity" of a telomerase reverse transcriptase (TRT). A modulator can be an agonist or an antagonist. A modulator can be any organic and inorganic compound; including, but not limited to, for example, small molecules, peptides, proteins, sugars,
30 nucleic acids, fatty acids and the like.

As used herein, "motif" refers to a sequence of contiguous amino acids

(or to a nucleic acid sequence that encodes a sequence of contiguous amino acids) that defines a feature or structure in a protein that is common to or conserved in all proteins of a defined class or type. The motif or consensus sequence may include both conserved and non-conserved residues. The conserved residues in the motif sequence indicate that the conserved residue or class (i.e., hydrophobic, polar, non-polar, or other class) of residues is typically present at the indicated location in each protein (or gene or mRNA) of the class of proteins defined by the motif. Motifs can differ in accordance with the class of proteins. Thus, for example, the reverse transcriptase enzymes form a class of proteins than can be defined by one or more motifs, and this class includes telomerase enzymes. However, the telomerase enzymes can also be defined as the class of enzymes with motifs characteristic for that class. Those of skill recognize that the identification of a residue as a conserved residue in a motif does not mean that every member of the class defined by the motif has the indicated residue (or class of residues) at the indicated position, and that one or more members of the class may have a different residue at the conserved position.

As used herein, the terms “**nucleic acid**” and “**polynucleotide**” are used interchangeably. Use of the term “polynucleotide” is not intended to exclude oligonucleotides (i.e., short polynucleotides) and can also refer to synthetic and/or non-naturally occurring nucleic acids (i.e., comprising nucleic acid analogues or modified backbone residues or linkages).

As used herein “**oligonucleotides**” or “**oligomers**” refer to a nucleic acid sequence of approximately 7 nucleotides or greater, and as many as approximately 100 nucleotides, which can be used as a primer, probe or amplifier. Oligonucleotides are often between about 10 and about 50 nucleotides in length, more often between about 14 and about 35 nucleotides, very often between about 15 and about 25 nucleotides, and the terms oligonucleotides or oligomers can also refer to synthetic and/or non-naturally occurring nucleic acids (i.e., comprising nucleic acid analogues or modified backbone residues or linkages).

As used herein, the term “**operably linked**,” refers to a functional relationship between two or more nucleic acid (e.g., DNA) segments: for example, a promoter or enhancer is operably linked to a coding sequence if it stimulates the

transcription of the sequence in an appropriate host cell or other expression system. Generally, sequences that are operably linked are contiguous, and in the case of a signal sequence both contiguous and in reading phase. However, enhancers need not be located in close proximity to the coding sequences whose transcription they enhance.

5 As used herein, the term "**polypeptide**" is used interchangeably herein with the term "protein," and refers to a polymer composed of amino acid residues linked by amide linkages, including synthetic, naturally-occurring and non-naturally occurring analogs thereof (amino acids and linkages). Peptides are examples of
10 polypeptides.

As used herein, a "**probe**" refers to a molecule that specifically binds another molecule. One example of a probe is a "**nucleic acid probe**" that specifically binds (i.e., anneals or hybridizes) to a substantially complementary nucleic acid. Another example of a probe is an "**antibody probe**" that specifically
15 binds to a corresponding antigen or epitope.

As used herein, "**recombinant**" refers to a polynucleotide synthesized or otherwise manipulated *in vitro* (e.g., "recombinant polynucleotide"), to methods of using recombinant polynucleotides to produce gene products in cells or other biological systems, or to a polypeptide ("recombinant protein") encoded by a
20 recombinant polynucleotide.

As used herein, a "**selection system**," in the context of stably transformed cell lines, refers to a method for identifying and/or selecting cells containing a recombinant nucleic acid of interest. A large variety of selection systems are known for identification of transformed cells and are suitable for use with
25 the present invention. For example, cells transformed by plasmids or other vectors can be selected by resistance to antibiotics conferred by genes contained on the plasmids, such as the well known amp, gpt, neo and hyg genes, or other genes such as the herpes simplex virus thymidine kinase (Wigler et al., Cell 11:223-32 [1977]) and adenine phosphoribosyltransferase (Lowy et al., Cell 22:817 [1980]) genes which
30 can be employed in tk- or ap^rt- cells, respectively. Also, antimetabolite, antibiotic or herbicide resistance can be used as the basis for selection; for example, *dhfr* which

confers resistance to methotrexate and is also useful for gene amplification (Wigler et al., *Proc. Natl. Acad. Sci.*, 77:3567 [1980]); *npt*, which confers resistance to the aminoglycosides neomycin and G-418 (Colbere-Garapin et al., *J. Mol. Biol.*, 150:1 [1981]) and *als* or *pat*, which confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively (Murry, in McGraw Hill Yearbook of Science and Technology, McGraw Hill, New York NY, pp 191-196, [1992]). Additional selectable genes have been described, for example, hygromycin resistance-conferring genes, *trpB*, which allows cells to utilize indole in place of tryptophan, or *hisD*, which allows cells to utilize histinol in place of histidine (Hartman and Mulligan, *Proc. Natl. Acad. Sci.*, 85:8047 [1988]). Recently, the use of visible markers has gained popularity with such markers as anthocyanins, beta-glucuronidase and its substrate, GUS, and luciferase and its substrate, luciferin, being widely used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system (Rhodes et al., *Meth. Mol. Biol.*, 55:121 [1995]).

As used herein, the "sequence" of a gene (unless specifically stated otherwise), nucleic acid, protein, or peptide refers to the order of nucleotides in either or both strands of a double-stranded DNA molecule, e.g., the sequence of both the coding strand and its complement, or of a single-stranded nucleic acid molecule, or to the order of amino acids in a peptide or protein.

As used herein, "specific binding" refers to the ability of one molecule, typically an antibody or polynucleotide, to contact and associate with another specific molecule even in the presence of many other diverse molecules. For example, a single-stranded polynucleotide can specifically bind to a single-stranded polynucleotide that is complementary in sequence, and an antibody specifically binds to (or "is specifically immunoreactive with") its corresponding antigen.

As used herein, "stringent hybridization conditions" or "stringency" refers to conditions in a range from about 5°C to about 20°C or 25°C below the melting temperature (T_m) of the target sequence and a probe with exact or nearly exact complementarity to the target. As used herein, the melting temperature is the temperature at which a population of double-stranded nucleic acid molecules

becomes half-dissociated into single strands. Methods for calculating the T_m of nucleic acids are well known in the art (see, e.g., Berger and Kimmel (1987) METHODS IN ENZYMOLOGY, VOL. 152: GUIDE TO MOLECULAR CLONING TECHNIQUES, San Diego: Academic Press, Inc. and Sambrook et al. (1989) MOLECULAR CLONING: A LABORATORY MANUAL, 2ND ED., VOLS. 1-3, Cold Spring Harbor Laboratory hereinafter, "Sambrook"), both incorporated herein by reference). As indicated by standard references, a simple estimate of the T_m value may be calculated by the equation: $T_m = 81.5 + 0.41(\% G + C)$, when a nucleic acid is in aqueous solution at 1 M NaCl (see e.g., Anderson and Young, *Quantitative Filter Hybridization in* NUCLEIC ACID HYBRIDIZATION (1985)). Other references include more sophisticated computations which take structural as well as sequence characteristics into account for the calculation of T_m . The melting temperature of a hybrid (and thus the conditions for stringent hybridization) is affected by various factors such as the length and nature (DNA, RNA, base composition) of the probe and nature of the target (DNA, RNA, base composition, present in solution or immobilized, and the like), and the concentration of salts and other components (e.g., the presence or absence of formamide, dextran sulfate, polyethylene glycol). The effects of these factors are well known and are discussed in standard references in the art, e.g., Sambrook, *supra* and Ausubel et al. *supra*. Typically, stringent hybridization conditions are salt concentrations less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion at pH 7.0 to 8.3, and temperatures at least about 30°C for short probes (e.g., 10 to 50 nucleotides) and at least about 60°C for long probes (e.g., greater than 50 nucleotides). As noted, stringent conditions may also be achieved with the addition of destabilizing agents such as formamide, in which case lower temperatures may be employed.

As used herein, the term "substantial identity," "substantial sequence identity," or "substantial similarity" in the context of nucleic acids, refers to a measure of sequence similarity between two polynucleotides. Substantial sequence identity can be determined by hybridization under stringent conditions, by direct comparison, or other means. For example, two polynucleotides can be identified as having substantial sequence identity if they are capable of specifically

hybridizing to each other under stringent hybridization conditions. Other degrees of sequence identity (e.g., less than "substantial") can be characterized by hybridization under different conditions of stringency. Alternatively, substantial sequence identity can be described as a percentage identity between two nucleotide (or polypeptide) sequences. Two sequences are considered substantially identical when they are at least about 60% identical, preferably at least about 70% identical, or at least about 80% identical, or at least about 90% identical, or at least about 95% or 98% to 100% identical. Percentage sequence (nucleotide or amino acid) identity is typically calculated by determining the optimal alignment between two sequences and comparing the two sequences. For example an exogenous transcript used for protein expression can be described as having a certain percentage of identity or similarity compared to a reference sequence (e.g., the corresponding endogenous sequence). Optimal alignment of sequences may be conducted using the local homology algorithm of Smith and Waterman (1981) *Adv. Appl. Math.* 2: 482, by the homology alignment algorithm of Needleman and Wunsch (1970) *J. Mol. Biol.* 48: 443, by the search for similarity method of Pearson and Lipman (1988) *Proc. Natl. Acad. Sci. U.S.A.* 85: 2444, by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by inspection. The best alignment (i.e., resulting in the highest percentage of identity) generated by the various methods is selected. Typically these algorithms compare the two sequences over a "comparison window" (usually at least 18 nucleotides in length) to identify and compare local regions of sequence similarity, thus allowing for small additions or deletions (i.e., gaps). Additions and deletions are typically 20 percent or less of the length of the sequence relative to the reference sequence, which does not comprise additions or deletions. It is sometimes desirable to describe sequence identity between two sequences in reference to a particular length or region (e.g., two sequences may be described as having at least 95% identity over a length of at least 500 basepairs). Usually the length will be at least about 50, 100, 200, 300, 400 or 500 basepairs, amino acids, or other residues. The percentage of sequence identity is calculated by comparing two optimally aligned sequences over the region of comparison,

determining the number of positions at which the identical nucleic acid base (e.g., A, T, C, G, or U) occurs in both sequences to yield the number of matched positions, and determining the number (or percentage) of matched positions as compared to the total number of bases in the reference sequence or region of comparison. An

5 additional algorithm that is suitable for determining sequence similarity is the BLAST algorithm, which is described in Altschul (1990) *J. Mol. Biol.* 215: 403-410; and Shpaer (1996) *Genomics* 38:179-191. Software for performing BLAST analyses is publicly available at the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>). This algorithm involves first identifying high

10 scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence that either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul *et al*, *supra*.). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs

15 containing them. The word hits are extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments;

20 or the end of either sequence is reached. The BLAST algorithm parameters W, T and X determine the sensitivity and speed of the alignment. The BLAST program uses as defaults a wordlength (W) of 11, the BLOSUM62 scoring matrix (*see* Henikoff (1992) *Proc. Natl. Acad. Sci. USA* 89: 10915-10919) alignments (B) of 50, expectation (E) of 10, M=5, N=-4, and a comparison of both strands. The term

25 BLAST refers to the BLAST algorithm which performs a statistical analysis of the similarity between two sequences; *see, e.g.*, Karlin (1993) *Proc. Natl. Acad. Sci. USA* 90:5873-5787. One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by

30 chance. For example, a nucleic acid can be considered similar to a TRT nucleic acid if the smallest sum probability in a comparison of the test nucleic acid to an TRT

nucleic acid is less than about 0.5, 0.2, 0.1, 0.01, or 0.001. Alternatively, another indication that two nucleic acid sequences are similar is that the polypeptide which the first nucleic acid encodes is immunologically cross reactive with the polypeptide encoded by the second nucleic acid.

5 As used herein, the terms "substantial identity," "substantial sequence identity," or "substantial similarity" in the context of a polypeptide, refers to a degree of similarity between two polypeptides in which a polypeptides comprises a sequence with at least 70% sequence identity to a reference sequence, or 80%, or 85% or up to 100% sequence identity to the reference sequence, or most preferably
10 90% identity over a comparison window of about 10-20 amino acid residues. Amino acid sequence similarity, or sequence identity, is determined by optimizing residue matches, if necessary, by introducing gaps as required. See Needleham et al. (1970) *J. Mol. Biol.* 48: 443-453; and Sankoff et al., 1983, *Time Warps, String Edits, and Macromolecules, The Theory and Practice of Sequence Comparison*, Chapter One,
15 Addison-Wesley, Reading, MA; and software packages from IntelliGenetics, Mountain View, CA, and the University of Wisconsin Genetics Computer Group, Madison, WI. As will be apparent to one of skill, the terms "substantial identity", "substantial similarity" and "substantial sequence identity" can be used interchangeably with regard to polypeptides or polynucleotides.

20 As used herein, the term "substantially pure," or "substantially purified," when referring to a composition comprising a specified reagent, such as an antibody (e.g. an anti-hTERT antibody), means that the specified reagent is at least about 75%, or at least about 90%, or at least about 95%, or at least about 99% or more of the composition (not including, e.g., solvent or buffer). Thus, for example,
25 a preferred immunoglobulin preparation of the invention that specifically binds an hTERT polypeptide is substantially purified.

 As used herein, a "telomerase negative" cell is one in which telomerase is not expressed, i.e., no telomerase catalytic activity can be detected using a conventional assay or a TRAP assay for telomerase catalytic activity. As
30 used herein, a "telomerase positive" cell is a cell in which telomerase is expressed (i.e. telomerase activity can be detected).

As used herein, a "telomerase-related" disease or condition is a disease or condition in a subject that is correlated with an abnormally high level of telomerase activity in cells of the individual, which can include any telomerase activity at all for most normal somatic cells, or which is correlated with a low level of telomerase activity that results in impairment of a normal cell function. Examples of telomerase-related conditions include, e.g., cancer (high telomerase activity in malignant cells) and infertility (low telomerase activity in germ-line cells).

As used herein, "test compound" or "agent" refers to any synthetic or natural compound or composition. The term includes all organic and inorganic compounds; including, for example, small molecules, peptides, proteins, sugars, nucleic acids, fatty acids and the like.

XII) EXAMPLES

The following examples are provided to illustrate the present invention and or to provide additional technical guidance not necessarily forming part of the invention, and not by way of limitation.

In the following sections, the following abbreviations apply: eq (equivalents); M (Molar); μ M (micromolar); N (Normal); mol (moles); mmol (millimoles); μ mol (micromoles); nmol (nanomoles); g (grams); mg (milligrams); μ g (micrograms); ng (nanograms); l or L (liters); ml (milliliters); μ l (microliters); cm (centimeters); mm (millimeters); μ m (micrometers); nm (nanometers); $^{\circ}$ C (degrees Centigrade); RPN (ribonucleoprotein); mreN (2'-O-methylribonucleotides); dNTP (deoxyribonucleotide); dH₂O (distilled water); DDT (dithiothreitol); PMSF (phenylmethylsulfonyl fluoride); TE (10 mM Tris HCl, 1 mM EDTA, approximately pH 7.2); KGlu (potassium glutamate); SSC (salt and sodium citrate buffer); SDS (sodium dodecyl sulfate); PAGE (polyacrylamide gel electrophoresis); Novex (Novex, San Diego, CA); BioRad (Bio-Rad Laboratories, Hercules, CA); Pharmacia (Pharmacia Biotech, Piscataway, NJ); Boehringer-Mannheim (Boehringer-Mannheim Corp., Concord, CA); Amersham (Amersham, Inc., Chicago, IL); Stratagene (Stratagene Cloning Systems, La Jolla, CA); NEB (New England Biolabs, Beverly,

MA); Pierce (Pierce Chemical Co., Rockford, IL); Beckman (Beckman Instruments, Fullerton, CA); Lab Industries (Lab Industries, Inc., Berkeley, CA); Eppendorf (Eppendorf Scientific, Madison, WI); and Molecular Dynamics (Molecular Dynamics, Sunnyvale, CA).

5

EXAMPLE 1

ISOLATION OF TELOMERASE PROTEINS AND CLONES

The following example details the isolation of telomerase proteins and clones from various organisms, including the euplates p. 123, hTRT, TRT and *S. pombe* TRT telomerase cDNA clones.

10

A. Background

i) Introduction

This section provides an overview of the purification and cloning of TRT genes, which is described in greater detail in subsequent sections of this Example. While telomerase RNA subunits have been identified in ciliates, yeast and mammals, protein subunits of the enzyme have not been identified as such prior to the present invention. Purification of telomerase from the ciliated protozoan *Euplotes aediculatus* yielded two proteins, termed p123 and p43 (see *infra*; Lingner (1996) *Proc. Natl. Acad. Sci. U.S.A.* 93:10712). *Euplotes aediculatus* is a hypotrichous ciliate having a macronucleus containing about 8×10^7 telomeres and about 3×10^5 molecules of telomerase. After purification, the active telomerase complex had a molecular mass of about 230 kD, corresponding to a 66 kD RNA subunit and two proteins of about 123 kD and 43 kD (Lingner (1996) *supra*). Photocross-linking experiments indicated that the larger p123 protein was involved in specific binding of the telomeric DNA substrate (Lingner, (1996) *supra*).

20

25

The p123 and p43 proteins were sequenced and the cDNA clones which encoded these proteins were isolated. These *Euplotes* sequences were found to be unrelated to the *Tetrahymena* telomerase-associated proteins p80 and p95. Sequence analysis of the *Euplotes* p123 revealed reverse transcriptase (RT) motifs. Furthermore, sequence analysis of the *Euplotes* p123 by comparison to other sequences revealed a

30

yeast homolog, termed Est2 protein (Lingner (1997) *Science* 276:561). Yeast Est2 had previously been shown to be essential for telomere maintenance *in vivo* (Lendvay (1996) *Genetics* 144:1399) but had not been identified as a telomerase catalytic protein. Site-specific mutagenesis demonstrated that the RT motifs of yeast Est2 are essential for
5 telomeric DNA synthesis *in vivo* and *in vitro* (Lingner (1997) *supra*).

ii) Identifying and Characterizing *S. pombe* Telomerase

PCR amplification of *S. pombe* DNA was carried out with degenerate sequence primers designed from the *Euplotes* p123 RT motifs as described below. Of
10 the four prominent PCR products generated, a 120 base pair band encoded a peptide sequence homologous to p123 and Est2. This PCR product was used as a probe in colony hybridization and identified two overlapping clones from an *S. pombe* genomic library and three from an *S. pombe* cDNA library. Sequence analysis revealed that none of the three *S. pombe* cDNA clones was full length, so RT-PCR was used to obtain the
15 sequences encoding the protein's N-terminus.

Complete sequencing of these clones revealed a putative *S. pombe* telomerase RT gene, *trt1*. The complete nucleotide sequence of *trt1* has been deposited in GenBank, accession number AF015783 (see Figure 15).

To test *S. pombe trt1* (as a catalytic subunit, two deletion constructs
20 were created. Analysis of the sequence showed that *trt1* encoded a basic protein with a predicted molecular mass of 116 kD. It was found that homology with p123 and Est2 was especially high in the seven reverse transcriptase motifs, underlined and designated as motifs 1, 2, A, B, C, D, and E (see Figure 63). An additional telomerase-specific motif, designated the T-motif, was also found. Fifteen introns, ranging in size from
25 36 to 71 base pairs, interrupted the coding sequence.

To test *S. pombe trt1* as a catalytic subunit, two deletion constructs were created. One removed only motifs B through D in the RT domains. The second removed 99% of the open reading frame.

Haploid cells grown from *S. pombe* spores of both mutants showed
30 progressive telomere shortening to the point where hybridization to telomeric repeats became almost undetectable. A *trt1*⁺/*trt1*⁻ diploid was sporulated and the resulting

tetrads were dissected and germinated on a yeast extract medium supplemented with amino acids (a YES plate, Alfa (1993) *Experiments with Fission Yeast*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY). Colonies derived from each spore were grown at 32°C for three days, and streaked successively to fresh YES plates every three days. A colony from each round was placed in six ml of YES liquid culture at 32°C and grown to stationary phase. Genomic DNA was prepared. After digestion with *ApaI*, DNA was subjected to electrophoresis on a 2.3% agarose gel, stained with ethidium bromide to confirm approximately equal loading in each lane, then transferred to a nylon membrane and hybridized to a telomeric DNA probe.

Senescence was indicated by the delayed onset of growth or failure to grow on agar (typically at the fourth streak-out after germination) and by colonies with increasingly ragged edges (colony morphology shown in Figure 22C) and by increasingly high fractions of elongated cells (as shown in Figure 22D). Cells were plated on Minimal Medium (Alfa (1993) *supra*) with glutamic acid substituted for ammonium chloride for two days at 32°C prior to photography.

When individual enlarged cells were separated on the dissecting microscope, the majority were found to undergo no further division. The same telomerase negative (*trt1⁻*) cell population always contained normal-sized cells which continued to divide, but which frequently produced non-dividing progeny. The telomerase-negative survivors may use a recombinational mode of telomere maintenance as documented in budding yeast strains that have various telomere-replication genes deleted (Lendvay (1996) *supra*, Lundblad (1993) *Cell* 73:347).

iii) Identifying and Characterizing Human Telomerase

An EST (expressed sequence tag) derived from human telomerase reverse transcriptase (hTRT) cDNA was identified by a BLAST search of the dbEST (expressed sequence tag) Genbank database using the *Euplotes* 123 kDa peptide and nucleic acid sequences, as well as the *Schizosaccharomyces* protein and corresponding cDNA (*tez1*) sequences. The EST, designated Genbank AA28196, is 389 nucleotides long and it corresponds to positions 1679 to 2076 of clone 712562 (Figure 18), was

obtained from the I.M.A.G.E. Consortium (Human Genome Center, DOE, Lawrence Livermore National Laboratory, Livermore, CA). This clone was obtained from a cDNA library of germinal B cells derived by flow sorting of tonsil cells. Complete sequencing of this hTRT cDNA clone showed all eight telomerase RT (TRT) motifs.

- 5 However, this hTRT clone did not encode a contiguous portion of a TRT because RT motifs B', C, D, and E, were contained in a different open reading frame than the more N-terminal RT motifs. In addition, the distance between RT motifs A and B was substantially shorter than that of the three previously known (non-human) TRTs.

- To isolate a full length cDNA clone, a cDNA library derived from the
10 human 293 cell line (described above) which expresses high levels of telomerase activity, was screened. A lambda cDNA library from the 293 cell line was partitioned into 25 pools containing about 200,000 plaques each. Each pool was screened by PCR with the primer pair 5'-CGGAAGAGTGTCTGGAGCAA-3' and 5'-GGATGAAGCGGAGTCTGGA-3'. Six subpools of one positive primary pool were
15 further screened by PCR using this same primer pair. For both the primary and the secondary subpool screening, hTRT was amplified for a total of 31 cycles at: 94°C, 45 seconds; 60°C, 45 seconds; and 72°C, 90 seconds. As a control, RNA of the house-keeping enzyme GAPDH was amplified using the primer pair 5'-CTCAGACACCA TGGGGAAGGTGA-3' and 5'-ATGATCTTGAGGCTGTTGTCATA-3' for a total of
20 16 cycles at 94°C, 45 seconds; 55°C, 45 seconds; and 72°C, 90 seconds.

- One hTRT positive subpool from the secondary screening was then screened by plaque hybridization with a probe from the 5' region of clone #712562. One phage was positively identified (designated Lambda phage 25-1.1, ATCC 209024, deposited May 12, 1997). It contained an approximately four kilobase insert, which was
25 excised and subcloned into the EcoRI site of pBluescript II SK+ vector (Stratagene, San Diego, CA) as an EcoRI fragment. This cDNA clone-containing plasmid was designated pGRN121. The cDNA insert totals approximately 4 kilobasepairs. The complete nucleotide sequence of the human hTRT cDNA (pGRN121) has been deposited in Genbank (accession AF015950) and the plasmid has been deposited with
30 the ATCC (ATCC 209016, deposited May 6, 1997).

B. Growth of *Euplotes aediculatus*

In this Example, cultures of *E. aediculatus* were obtained from Dr. David Prescott, MCDB, University of Colorado. Dr. Prescott originally isolated this culture from pond water, although this organism is also available from the ATCC (ATCC #30859). Cultures were grown as described by Swanton *et al.*, (Swanton *et al.*, Chromosoma 77:203 [1980]), under non-sterile conditions, in 15-liter glass containers containing *Chlorogonium* as a food source. Organisms were harvested from the cultures when the density reached approximately 10^4 cells/ml.

10

C. Preparation of Nuclear Extracts

In this Example, nuclear extracts of *E. aediculatus* were prepared using the method of Lingner *et al.*, (Lingner *et al.*, Genes Develop., 8:1984 [1994]), with minor modifications, as indicated below. Briefly, cells grown as described in Part B were concentrated with 15 μ m Nytex filters and cooled on ice. The cell pellet was resuspended in a final volume of 110 ml TMS/PMSF/spermidine phosphate buffer. The stock TMS/PMSF/spermidine phosphate buffer was prepared by adding 0.075 g spermidine phosphate (USB) and 0.75 ml PMSF (from 100 mM stock prepared in ethanol) to 150 ml TMS. TMS comprised 10 mM Tris-acetate, 10 mM $MgCl_2$, 85.5752 g sucrose/liter, and 0.33297 g $CaCl_2$ /liter, pH 7.5.

After resuspension in TMS/PMSF/spermidine phosphate buffer, 8.8 ml 10% NP-40 and 94.1 g sucrose were added and the mixture placed in a siliconized glass beaker with a stainless steel stirring rod attached to an overhead motor. The mixture was stirred until the cells were completely lysed (approximately 20 minutes). The mixture was then centrifuged for 10 minutes at 7500 rpm (8950 x g), at 4°C, using a Beckman JS-13 swing-out rotor. The supernatant was removed and nuclei pellet was resuspended in TMS/PMSF/spermidine phosphate buffer, and centrifuged again, for 5 minutes at 7500 rpm (8950 x g), at 4°C, using a Beckman (RTM) JS-13 swing-out rotor.

The supernatant was removed and the nuclei pellet was resuspended in a buffer comprised of 50 mM Tris-acetate, 10 mM $MgCl_2$, 10% glycerol, 0.1% NP-

40, 0.4 M KGlu, 0.5 mM PMSF, pH 7.5, at a volume of 0.5 ml buffer per 10 g of harvested cells. The resuspended nuclei were then dounced in a glass homogenizer with approximately 50 strokes, and then centrifuged for 25 minutes at 14,000 rpm at 4°C, in an Eppendorf (RTM) centrifuge. The supernatant containing the nuclear
5 extract was collected, frozen in liquid nitrogen, and stored at -80°C until used.

D. Purification of Telomerase

In this Example, nuclear extracts prepared as described in Part C were used to purify *E. aedicularis* telomerase. In this purification protocol, telomerase was first
10 enriched by chromatography on an Affi-Gel(RTM)-heparin column, and then extensively purified by affinity purification with an antisense oligonucleotide. As the template region of telomerase RNA is accessible to hybridization in the telomerase RNP particle, an antisense oligonucleotide (*i.e.*, the "affinity oligonucleotide") was synthesized that was complementary to this template region as an affinity bait for the
15 telomerase. A biotin residue was included at the 5' end of the oligonucleotide to immobilize it to an avidin column.

Following the binding of the telomerase to the oligonucleotide, and extensive washing, the telomerase was eluted by use of a displacement oligonucleotide. The affinity oligonucleotide included DNA bases that were not
20 complementary to the telomerase RNA 5' to the telomerase-specific sequence. As the displacement oligonucleotide was complementary to the affinity oligonucleotide for its entire length, it was able to form a more thermodynamically stable duplex than the telomerase bound to the affinity oligonucleotide. Thus, addition of the displacement oligonucleotide resulted in the elution of the telomerase from the
25 column.

The nuclear extracts prepared from 45 liter cultures were frozen until a total of 34 ml of nuclear extract was collected. This corresponded to 630 liters of culture (*i.e.*, approximately 4×10^9 cells). The nuclear extract was diluted with a buffer to 410 ml, to provide final concentrations of 20 mM Tris-acetate, 1 mM
30 $MgCl_2$, 0.1 mM EDTA, 33 mM KGlu, 10% (vol/vol) glycerol, 1 mM dithiothreitol (DTT), and 0.5 mM-phenylmethylsulfonyl-fluoride (PMSF), at a pH of 7.5.

The diluted nuclear extract was applied to an Affi-Gel(RTM)-heparin gel column (Bio-Rad), with a 230 ml bed volume and 5 cm diameter, equilibrated in the same buffer and eluted with a 2-liter gradient from 33 to 450 mM KGlu. The column was run at 4°C, at a flow rate of 1 column volume/hour. Fractions of 50 mls each were collected and assayed for telomerase activity as described in Part E. Telomerase was eluted from the column at approximately 170 mM KGlu. Fractions containing telomerase (approximately 440 ml) were pooled and adjusted to 20 mM Tris-acetate, 10 mM MgCl₂, 1 mM EDTA, 300 mM KGlu, 10% glycerol, 1 mM DTT, and 1% Nonidet (RTM) P-40. This buffer was designated as "WB."

To this preparation, 1.5 nmol of each of two competitor DNA oligonucleotides (5'-TAGACCTGTTAGTGTACATTTGAATTGAAGC-3' (and (5'-TAGACCTGTTAGGTTGGATTGTGGCATCA-3', 50 µg yeast RNA (Sigma), and 0.3 nmol of biotin-labeled telomerase-specific oligonucleotide (5'-biotin-TAGACCTGTTA-(mreG)₂-(rmeU)₄-(rmeG)₄-(rmeU)₄-remG-3'), were added per ml of the pool. The 2-*O*-methyribonucleotides of the telomerase specific oligonucleotides were complementary to the the telomerase RNA; template region; the deoxyribonucleotides were not complementary. The inclusion of competitor, non-specific DNA oligonucleotides increased the efficiency of the purification, as the effects of nucleic acid binding proteins and other components in the mixture that would either bind to the affinity oligonucleotide or remove the telomerase from the mixture were minimized.

This material was then added to Ultralink immobilized neutravidin plus (Pierce) column material, at a volume of 60 µl of suspension per ml of pool. The column material was pre-blocked twice for 15 minutes each blocking, with a preparation of WB containing 0.01% Nonidet P-40, 0.5 mg BSA, 0.5 mg/ml lysozyme, 0.05 mg/ml glycogen, and 0.1 mg/ml yeast RNA. The blocking was conducted at 4°C, using a rotating wheel to block the column material thoroughly. After the first blocking step, and before the second blocking step, the column material was centrifuged at 200 x g for 2 minutes to pellet the matrix.

The pool-column mixture was incubated for 8 minutes at 30°C, and

then for an additional 2 hours at 4°C, on a rotating wheel (approximately 10 rpm; Labindustries) to allow binding. The pool-column mixture was then centrifuged 200 xg for 2 minutes, and the supernatant containing unbound material was removed. The pool-column mixture was then washed. This washing process included the steps of rinsing the pool-column mixture with WB at 4°C, washing the mixture for 15 minutes with WB at 4°C, rinsing with WB, washing for 5 minutes at 30°C, with WB containing 0.6 M KGlu, and no Nonidet P-40, washing 5 minutes at 25°C with WB, and finally, rinsing again with WB. The volume remaining after the final wash was kept small, in order to yield a ratio of buffer to column material of approximately 1:1.

Telomerase was eluted from the column material by adding 1 nmol of displacement deoxyoligonucleotide (5'-CA₄C₄A₄C₂TA₂CAG₂TCTA-3'), per ml of column material and incubating at 25°C for 30 minutes. The material was centrifuged for 2 minutes at 14,000 rpm in a microcentrifuge (Eppendorf), and the eluate collected. The elution procedure was repeated twice more, using fresh displacement oligonucleotide each time. As mentioned above, because the displacement oligonucleotide was complementary to the affinity oligonucleotide, it formed a more thermodynamically stable complex with the affinity oligonucleotide than P-40. Thus, addition of the displacement oligonucleotide to an affinity-bound telomerase resulted in efficient elution of telomerase under native conditions. The telomerase appeared to be approximately 50% pure at this stage, as judged by analysis on a protein gel. The affinity purification of telomerase and elution with a displacement oligonucleotide is shown in Figure 26 (panels A and B, respectively). In this Figure, the 2'-O-methyl sugars of the affinity oligonucleotide are indicated by the bold line. The black and shaded oval shapes in this Figure are intended to represent graphically the protein subunits of the present invention.

The protein concentrations of the extract and material obtained following Affi-Gel-heparin column chromatography were determined using the method of Bradford (Bradford, Anal. Biochem., 72:248 [1976]), using BSA as the standard. Only a fraction of the telomerase preparation was further purified on a

glycerol gradient.

The sedimentation coefficient of telomerase was determined by glycerol gradient centrifugation, as described in Part I.

Table 5 below is a purification table for telomerase purified according to the methods of this Example. The telomerase was enriched 12-fold in nuclear extracts, as compared to whole cell extracts, with a recovery of 80%; 85% of telomerase was solubilized from nuclei upon extraction.

Table 5. Purification of Telomerase

Fraction	Protein (mg)	Telomerase (pmol of RNP)	Telomerase/ Protein/pmol of RNP/mg	Recovery (%)	Purification Factor
Nuclear Extract	2020	1720	0.9	100	1
Heparin	125	1040	8.3	60	10
Affinity	0.3**	680	2270	40	2670
Glycerol Gradient	NA*	NA*	NA*	25	NA*

*NA=Not available

10 **This value was calculated from the measured amount of telomerase (680 pmol), by assuming a purity of 50% (based on a protein gel).

E. Telomerase Activity

At each step in the purification of telomerase, the preparation was analyzed by three separate assays, one of which was activity, as described in this Example. In general, telomerase assays were done in 40 μ l containing 0.003-0.3 μ l of nuclear extract, 50 mM Tris-Cl (pH 7.5), 50 mM KGlu, 10 mM $MgCl_2$, 1 mM DTT, 125 μ M dTTP, 125 μ M dGTP, and approximately 0.2 pmoles of 5'- ^{32}P -labelled oligonucleotide substrate (*i.e.*, approximately 400,000 cpm). Oligonucleotide primers were heat-denatured prior to their addition to the reaction mixture. Reactions were assembled on ice and incubated for 30 minutes at 25°C. The reactions were stopped by addition of 200 μ l of 10 mM Tris-Cl (pH 7.5), 15 mM EDTA, 0.6% SDS, and 0.05 mg/ml proteinase K, and incubated for at least 30 minutes at 45°C. After ethanol precipitation, the products were analyzed on denaturing 8% PAGE gels, as known in the art (*See e.g.*, Sambrook *et al.*, 1989).

F. Quantitation of Telomerase Activity

In this Example, quantitation of telomerase activity through the purification procedure is described. Quantitation was accomplished by assaying the elongation of oligonucleotide primers in the presence of dGTP and [α - ^{32}P]dTTP. Briefly, 1 μ M 5'-(G₄T₄)₂-3' oligonucleotide was extended in a 20 μ l reaction mixture in the presence of 2 μ l of [α - ^{32}P]dTTP (10 mCi/ml, 400 Ci/mmol; 1 Ci=37 GBq), and 125 μ M dGTP as described (Lingner *et al.*, Genes Develop., 8:1984 [1994]) and loaded onto an 8% PAGE sequencing gel as described.

The results of this study are shown in Figure 28. In lane 1, there is no telomerase present (*i.e.*, a negative control); lanes 2, 5, 8, and 11 contained 0.14 fmol telomerase; lanes 3, 6, 9, and 12 contained 0.42 fmol telomerase; and lanes 4, 7, 10, and 13 contained 1.3 fmol telomerase. Activity was quantitation using a PhosphorImager (Molecular Dynamics) using the manufacturer's instructions. It was determined that under these conditions, 1 fmol of affinity-purified telomerase incorporated 21 fmol of dTTP in 30 minutes.

As shown in Figure 28, the specific activity of the telomerase did not change significantly through the purification procedure. Affinity-purified telomerase

was fully active. However, it was determined that at high concentrations, an inhibitory activity was detected and the activity of crude extracts was not linear. Thus, in the assay shown in Figure 28, the crude extract was diluted 700-7000-fold. Upon purification, this inhibitory activity was removed and no inhibitory effect was
5 detected in the purified telomerase preparations, even at high enzyme concentrations.

G. Gel Electrophoresis and Northern Blots

As stated in Part E, at each step in the purification of telomerase, the preparation was analyzed by three separate assays. This Example describes the gel
10 electrophoresis and blotting procedures used to quantify telomerase RNA present in fractions and analyze the integrity of the telomerase ribonucleoprotein particle.

i) Denaturing Gels and Northern Blots

In this Example, synthetic T7-transcribed telomerase RNA of known
15 concentration served as the standard. Throughout this investigation, the RNA component was used as a measure of telomerase.

A construct for phage T7 RNA polymerase transcription of *E. aediculatus* telomerase RNA was produced, using (PCR). The telomerase RNA gene was amplified with primers that annealed to either end of the gene. The primer that
20 annealed at the 5' end also encoded a hammerhead ribozyme sequence to generate the natural 5' end upon cleavage of the transcribed RNA, a T7-promoter sequence, and an *EcoRI* site for subcloning. The sequence of this 5' primer was 5'-
GCGGGAATTCTAA
TACGACTCACTATAGGGAAGAACTCTGATGAGGCCGAAAGGCCGAAACT
25 CCACGAAAGTGGAGTAAGTTTCTCGATAATTGATCTGTAG-3'. The 3' primer included an *EarI* site for termination of transcription at the natural 3' end, and a *BamHI* site for cloning. The sequence of this 3' primer was 5'-
CGGGGATCCTCTTCAAAG
ATGAGAGGACAGCAAAC-3'. The PCR amplification product was cleaved with
30 *EcoRI* and *BamHI*, and subcloned into the respective sites of pUC19 (NEB), to give "pEaT7." The correctness of this insert was confirmed by DNA sequencing. T7

transcription was performed as described by Zaug *et al.*, Biochemistry 33:14935 [1994], with *Eco*I-linearized plasmid. RNA was gel-purified and the concentration was determined (an A_{260} of 1 = 40 μ g/ml). This RNA was used as a standard to determine the telomerase RNA present in various preparations of telomerase.

5 The signal of hybridization was proportional to the amount of telomerase RNA, and the derived RNA concentrations were consistent with, but slightly higher than those obtained by native gel electrophoresis. Comparison of the amount of whole telomerase RNA in whole cell RNA to serial dilutions of known T7 RNA transcript concentrations indicated that each *E. aediculatus* cell contained
10 approximately 300,000 telomerase molecules.

 Visualization of the telomerase was accomplished by Northern blot hybridization to its RNA component, using methods as described (Linger *et al.*, Genes Develop., 8:1984 [1994]). Briefly, RNA (less than or equal to 0.5 μ g/lane) was resolved on an 8% PAGE and electroblotted onto a Hybond(RTM)-N membrane
15 (Amersham), as known in the art (*see e.g.*, Sambrook *et al.*, 1989). The blot was hybridized overnight in 10 ml of 4x SSC, 10x Denhardt's solution, 0.1% SDS, and 50 μ g/ml denatured herring sperm DNA. After pre-hybridizing for 3 hours, 2×10^6 cpm probe/ml hybridization solution was added. The randomly labelled probe was a PCR-product that covered the entire telomerase RNA gene. The blot was washed
20 with several buffer changes for 30 minutes in 2x SSC, 0.1% SDS, and then washed for 1 hour in 0.1x SSC and 0.1% SDS at 45°C.

ii) Native Gels and Northern Blots

 In this experiment, the purified telomerase preparation was run on
25 native (*i.e.*, non-denaturing) gels of 3.5% polyacrylamide and 0.33% agarose, as known in the art and described (Lamond and Sproat, [1994], *supra*). The telomerase comigrated approximately with the xylene cyanol dye.

 The native gel results indicated that telomerase was maintained as an RNP throughout the purification protocol. Figure 27 is a photograph of a Northern
30 blot showing the mobility of the telomerase in different fractions on a non-denaturing

gel as well as *in vitro* transcribed telomerase. In this figure, lane 1 contained 1.5 fmol telomerase RNA, lane 2 contained 4.6 fmol telomerase RNA, lane 3 contained 14 fmol telomerase RNA, lane 4 contained 41 fmol telomerase RNA, lane 5 contained nuclear extract (42 fmol telomerase), lane 6 contained Affi-Gel-heparin-purified telomerase (47 fmol telomerase), lane 7 contained affinity-purified telomerase (68 fmol), and lane 8 contained glycerol gradient-purified telomerase (35 fmol).

As shown in Figure 27, in nuclear extracts, the telomerase was assembled into an RNP particle that migrated slower than unassembled telomerase RNA. Less than 1% free RNA was detected by this method. However, a slower migrating telomerase RNP complex was also sometimes detected in extracts. Upon purification on the Affi-Gel-heparin column, the telomerase RNP particle did not change in mobility (Figure 27, lane 6). However, upon affinity purification the mobility of the RNA particle slightly increased (Figure 27, lane 7), perhaps indicating that a protein subunit or fragment had been lost. On glycerol gradients, the affinity-purified telomerase did not change in size, but approximately 2% free telomerase RNA was detectable (Figure 27, lane 8), suggesting that a small amount of disassembly of the RNP particle had occurred.

20 H. Telomerase Protein Composition

In this Example, the analysis of the purified telomerase protein composition are described.

Glycerol gradient fractions obtained as described in Part D, were separated on a 4-20% polyacrylamide gel (Novex). Following electrophoresis, the gel was stained with Coomassie (RTM) brilliant blue. Figure 29 shows a photograph of the gel. Lanes 1 and 2 contained molecular mass markers (Pharmacia) as indicated on the left side of the gel shown in Figure 29. Lanes 3-5 contained glycerol gradient fraction pools as indicated on the top of the gel (*i.e.*, lane 3 contained fractions 9-14, lane 4 contained fractions 15-22, and lane 5 contained fractions 23-32). Lane 4 contained the pool with 1 pmol of telomerase RNA. In lanes 6-9 BSA standards were run at concentrations indicated at the top of the gel in Figure 29 (*i.e.*, lane 6

contained 0.5 pmol BSA, lane 7 contained 1.5 pmol BSA, lane 8 contained 4.5 BSA, and lane 9 contained 15 pmol BSA).

As shown in Figure 29, polypeptides with molecular masses of 120 and 43 kDa co-purified with the telomerase. The 43 kDa polypeptide was observed as a doublet. It was noted that the polypeptide of approximately 43 kDa in lane 3 migrated differently than the doublet in lane 4; it may be an unrelated protein. The 120 kDa and 43 kDa doublet each stained with Coomassie brilliant blue at approximately the level of 1 pmol, when compared with BSA standards. Because this fraction contained 1 pmol of telomerase RNA, all of which was assembled into an RNP particle (See, Figure 27, lane 8), there appear to be two polypeptide subunits that are stoichiometric with the telomerase RNA. However, it is also possible that the two proteins around 43 kDa are separate enzyme subunits.

Affinity-purified telomerase that was not subjected to fractionation on a glycerol gradient contained additional polypeptides with apparent molecular masses of 35 and 37 kDa, respectively. This latter fraction was estimated to be at least 50% pure. However, the 35 kDa and 37 kDa polypeptides that were present in the affinity-purified material were not reproducibly separated by glycerol gradient centrifugation. These polypeptides may be contaminants, as they were not visible in all activity-containing preparations.

I. Sedimentation Coefficient

The sedimentation coefficient for telomerase was determined by glycerol gradient centrifugation. In this Example, nuclear extract and affinity-purified telomerase were fractionated on 15-40% glycerol gradients containing 20 mM Tris-acetate, with 1 mM $MgCl_2$, 0.1 mM EDTA, 300 mM KGlu, and 1 mM DTT, at pH 7.5. Glycerol gradients were poured in 5 ml (13 x 51 mm) tubes, and centrifuged using an SW55Ti rotor (Beckman) at 55,000 rpm for 14 hours at 4°C.

Marker proteins were run in a parallel gradient and had a sedimentation coefficient of 7.6 S for alcohol dehydrogenase (ADH), 113 S for catalase, 17.3 S for apoferritin, and 19.3 S for thyroglobulin. The telomerase peak was identified by native gel electrophoresis of gradient fractions followed by blot

hybridization to its RNA component.

Figure 30 is a graph showing the sedimentation coefficient for telomerase. As shown in this Figure, affinity-purified telomerase co-sedimented with catalase at 11.5 S, while telomerase in nuclear extracts sedimented slightly faster, peaking around 12.5 S. Therefore, consistent with the mobility of the enzyme in native gels, purified telomerase appears to have lost a proteolytic fragment or a loosely associated subunit.

The calculated molecular mass for telomerase, if it is assumed to consist of one 120 kDa protein subunit, one 43 kDa subunit, and one RNA subunit of 66 kDa, adds up to a total of 229 kDa. This is in close agreement with the 232 kDa molecular mass of catalase. However, the sedimentation coefficient is a function of the molecular mass, as well as the partial specific volume and the frictional coefficient of the molecule, both of which are unknown for the *Euplotes* telomerase RNP.

J. Substrate Utilization

In this Example, the substrate requirements of *Euplotes* telomerase were investigated. One simple model for DNA end replication predicts that after semi-conservative DNA replication, telomerase extends double-stranded, blunt-ended DNA molecules. In a variation of this model, a single-stranded 3' end is created by a helicase or nuclease after replication. This 3' end is then used by telomerase for binding and extension.

To determine whether telomerase is capable of elongating blunt-ended molecules, model hairpins were synthesized with telomeric repeats positioned at their 3' ends. These primer substrates were gel-purified, 5'-end labelled with polynucleotide kinase, heated at 0.4 μ M to 80°C for 5 minutes, and then slowly cooled to room temperature in a heating block, to allow renaturation and helix formation of the hairpins. Substrate mobility on a non-denaturing gel indicated that very efficient hairpin formation was present, as compared to dimerization.

Assays were performed with unlabelled 125 μ M dGTP, 125 μ M dTTP, and 0.02 μ M 5'-end-labelled primer (5'-³²P-labelled oligonucleotide substrate)

in 10 µl reaction mixtures that contained 20 mM Tris-acetate, with 10 mM MgCl₂, 50 mM KGlu, and 1 mM DTT, at pH 7.5. These mixtures were incubated at 25°C for 30 minutes. Reactions were stopped by adding formamide loading buffer (*i.e.*, TBE, formamide, bromthymol blue, and cyanol, Sambrook, 1989, *supra*).

- 5 Primers were incubated without telomerase (" - "), with 5.9 fmol of affinity-purified telomerase (" + "), or with 17.6 fmol of affinity-purified telomerase (" + + + "). Affinity-purified telomerase used in this assay was dialyzed with a membrane having a molecular cut-off of 100 kDa, in order to remove the displacement oligonucleotide. Reaction products were separated on an 8%
10 PAGE/urea gel containing 36% formamide, to denature the hairpins. The sequences of the primers used in this study, as well as their lane assignments are shown in Table 6.

TABLE 6. Primer Sequences

Lane	Primer Sequence (5' to 3')
1-3	C ₄ (A ₄ C ₄) ₃ CACA(G ₄ T ₄) ₃ G ₄
4-6	C ₂ (A ₄ C ₄) ₃ CACA(G ₄ T ₄) ₃ G ₄
7-9	(A ₄ C ₄) ₃ CACA(G ₄ T ₄) ₃ G ₄
10-12	A ₂ C ₄ (A ₄ C ₄) ₂ CACA(G ₄ T ₄) ₃ G ₄
13-15	C ₄ (A ₄ C ₄) ₂ CACA(G ₄ T ₄) ₃
16-18	(A ₄ C ₄) ₃ CACA(G ₄ T ₄) ₃
19-21	A ₂ C ₄ (A ₄ C ₄) ₂ CACA(G ₄ T ₄) ₃
22-24	C ₄ (A ₄ C ₄) ₂ CACA(G ₄ T ₄) ₃
25-27	C ₂ (A ₄ C ₄) ₂ CACA(G ₄ T ₄) ₃
28-30	(A ₄ C ₄) ₂ CACA(G ₄ T ₄) ₃

The gel results are shown in Figure 31. Lanes 1-15 contained substrates with telomeric repeats ending with four G residues. Lanes 16-30 contained substrates with telomeric repeats ending with four T residues. The putative alignment on the telomerase RNA template is indicated in Figure 32. It was assumed that the primer sets anneal at two very different positions in the template shown in Figure 32 (i.e., Panel A and Panel B, respectively). This may have affected their binding and/or elongation rate.

Figure 33 shows a lighter exposure of lanes 25-30 in Figure 31. The lighter exposure of Figure 33 was taken to permit visualization of the nucleotides that are added and the positions of pausing in elongated products. Percent of substrate elongated for the third lane in each set was quantified on a PhosphorImager, as indicated on the bottom of Figure 31.

The substrate efficiencies for these hairpins were compared with double-stranded telomere-like substrates with overhangs of differing lengths. A model substrate that ended with four G residues (see lanes 1-15 of Figure 31) was not elongated when it was blunt ended (see lanes 1-3). However, slight extension was observed with an overhang length of two bases; elongation became efficient when the overhang was at least 4 bases in length. The telomerase acted in a similar manner with a double-stranded substrate that ended with four T residues, with a 6-base overhang required for highly efficient elongation. In Figure 31, the faint bands below the primers in lanes 10-15 that are independent of telomerase represent shorter oligonucleotides in the primer preparations.

The lighter exposure of lanes 25-30 in Figure 33 shows a ladder of elongated products, with the darkest bands correlating with the putative 5' boundary of the template (as described by Lingner *et al.*, Genes Develop., 8:1984 [1994]). The abundance of products that correspond to other positions in the template suggested that pausing and/or dissociation occurs at sites other than the site of translocation with the purified telomerase.

As shown in Figure 31, double-stranded, blunt-ended oligonucleotides were not substrates for telomerase. To determine whether these molecules would

bind to telomerase, a competition experiment was performed. In this experiment, 2 nM of 5'-end labeled substrate with the sequence (G₄T₄)₂, or a hairpin substrate with a six base overhang were extended with 0.125 nM telomerase (Figure 31, lanes 25-27). Although the same unlabeled oligonucleotide substrates competed efficiently with labeled substrate for extension, no reduction of activity was observed when the double-stranded blunt-ended hairpin oligonucleotides were used as competitors, even in the presence of 100-fold excess hairpins.

These results indicated that double-stranded, blunt-ended oligonucleotides cannot bind to telomerase at the concentrations and conditions tested in this Example. Rather, a single-stranded 3' end is required for binding. It is likely that this 3' end is required to base pair with the telomerase RNA template.

K. Cloning & Sequencing of the 123 kDa Polypeptide

In this Example, the cloning of the 123 kDa polypeptide of Euplotes telomerase (*i.e.*, the 123 kDa protein subunit) is described. In this study, an internal fragment of the telomerase gene was amplified by PCR, with oligonucleotide primers designed to match peptide sequences that were obtained from the purified polypeptide obtained in Part D, above. The polypeptide sequence was determined using the nanoES tandem mass spectroscopy methods known in the art and described by Calvio *et al.*, RNA 1:724-733 [1995]. The oligonucleotide primers used in this Example had the following sequences, with positions that were degenerate shown in parentheses--5'-TCT(G/A)

AA(G/A)TA(G/A)TG(T/G/A)GT(G/A/T/C)A(T/G/A)(G/A)TT(G/A)TTCAT-3', and 5'-GCGGATCCATGAA(T/C)CC(A/T)GA(G/A)AA(T/C)CC(A/T)AA(T/C)GT-3'.

A 50 µl reaction contained 0.2 mM dNTPs, 0.15 µg *E. aediculatus* chromosomal DNA, 0.5 µl *Taq* (Boehringer-Mannheim), 0.8 µg of each primer, and 1x reaction buffer (Boehringer-Mannheim). The reaction was incubated in a thermocycler (Perkin-Elmer), using the following--5 minutes at 95°C, followed by 30 cycles of 1 minute at 94°C, 1 minute at 52°C, and 2 minutes at 72°C. The reaction was completed by a 10 minute incubation at 72°C.

A genomic DNA library was prepared from the chromosomal *E. aediculatus* DNA by cloning blunt-ended DNA into the *Sma*I site of pCR-Script plasmid vector Figure 14(Stratagene). This library was screened by colony hybridization, with the radiolabelled, gel-purified PCR product. Plasmid DNA of positive clones was prepared and sequenced by the dideoxy method (Sanger *et al.*, Proc. Natl. Acad. Sci., 74:5463 [1977]) or manually, through use of an automated sequencer (ABI). The DNA sequence of the gene encoding this polypeptide is shown in Figure 13. The start codon in this sequence inferred from the DNA sequence, is located at nucleotide position 101, and the open reading frame ends at position 3193.

The genetic code of *Euplotes* differs from other organisms in that the "UGA" codon encodes a cysteine residue. The amino acid sequence of the polypeptide inferred from the DNA sequence is shown in Figure 14, and assumes that no unusual amino acids are inserted during translation and no post-translational modification occurs.

15 L. Cloning & Sequencing of the 43 kDa Polypeptide

In this Example, the cloning of the 43 kDa polypeptide of telomerase (*i.e.*, the 43 kDa protein subunit) is described. In this study, an internal fragment of the corresponding telomerase gene was amplified by PCR, with oligonucleotide primers designed to match peptide sequences that were obtained from the purified polypeptide obtained in Part D, above. The polypeptide sequence was determined using the nanoES tandem mass spectroscopy methods known in the art and described by Calvio *et al.*, *supra*. The oligonucleotide primers used in this Example had the following sequences--5'-NNNGTNAC(C/T/A)GG(C/T/A)AT(C/T/A)AA(C/T)AA-3', and 5'-(T/G/A)GC

(T/G/A)GT(C/T)TC(T/C)TG(G/A)TC(G/A)TT(G/A)TA-3'. In this sequence, "N" indicates the presence of any of the four nucleotides (*i.e.*, A, T, G, or C).

The PCR was performed as described in Part K.

A genomic DNA library was prepared and screened as described in Part K. The DNA sequence of the gene encoding this polypeptide is shown in Figure 34. Three potential reading frames are shown for this sequence, as shown in Figure 35. For clarity, the amino acid sequence is indicated below the nucleotide sequence

in all three reading frames. These reading frames are designated as "a," "b," and "c". A possible start codon is encoded at nucleotide position 84 in reading frame "c." The coding region could end at position 1501 in reading frame "b." Early stop codons, indicated by asterisks in this figure, occur in all three reading frames
5 between nucleotide position 337-350.

The "La-domain" is indicated in bold-face type. Further downstream, the protein sequence appears to be encoded by different reading frames, as none of the three frames is uninterrupted by stop codons. Furthermore, peptide sequences from purified protein are encoded in all three frames. Therefore, this gene appears to
10 contain intervening sequences, or in the alternative, the RNA is edited. Other possibilities include ribosomal frame-shifting or sequence errors. However, the homology to the La-protein sequence remains of significant interest. Again, in *Euplotes*, the "UGA" codon encodes a cysteine residue.

15 M. Amino Acid and Nucleic Acid Comparisons

In this Example, comparisons between various reported sequences and the sequences of the 123 kDa and 43 kDa telomerase subunit polypeptides were made.

20 i) Comparisons with the 123 kDa *E. aediculatus* Telomerase Subunit

The amino acid sequence of the 123 kDa *Euplotes aediculatus* polypeptide was compared with the sequence of the 80 kDa telomerase protein subunit of *Tetrahymena thermophila* (GenBank accession #U25641) to investigate
25 their similarity. The nucleotide sequence as obtained from GenBank encoding this protein is shown in Figure 42. The amino acid sequence of this protein as obtained from GenBank is shown in Figure 43. The sequence comparison between the 123 kDa *E. aediculatus* and 80 kDa *T. thermophila* is shown in Figure 36. In this figure, the *E. aediculatus* sequence is the upper sequence, while the *T. thermophila* sequence
30 is the lower sequence. The observed identity was determined to be approximately 19%, while the percent similarity was approximately 45%, values similar to what

would be observed with any random protein sequence. In Figures 36-39, identities are indicated by vertical bars, while single dots between the sequences indicate somewhat similar amino acids, and double dots between the sequences indicate more similar amino acids.

The amino acid sequence of the 123 kDa *Euploes*
5 *aediculatus* polypeptide was also compared with the sequence of the 95 kDa telomerase protein subunit of *Tetrahymena thermophila* (GenBank accession #U25642), to investigate their similarity. The nucleotide sequence as obtained from GenBank encoding this protein is shown in Figure 44. The amino acid sequence of this protein as obtained from GenBank is shown in Figure 45. This sequence
10 comparison is shown in Figure 37. In this figure, the *E. aediculatus* sequence is the upper sequence), while the *T. thermophila* sequence is the lower sequence. The observed identity was determined to be approximately 20%, while the percent similarity was approximately 43%, values similar to what would be observed with any random protein sequence.

15 Significantly, the amino acid sequence of the 123 kDa *E. aediculatus* polypeptide contains the five motifs characteristic of reverse transcriptases. The 123 kDa polypeptide was also compared with the polymerase domains of various reverse transcriptases. Figure 40 shows the alignment of the 123 kDa polypeptide with the putative yeast homolog (L8543.12 or ESTp). The amino acid sequence of L8543.12
20 obtained from GenBank is shown in Figure 46.

Four motifs (A, B, C, and D) were included in this comparison. In this Figure 40, highly conserved residues are indicated by white letters on a black background. Residues of the *E. aediculatus* sequences that are conserved in the other sequence are indicated in bold; the "h" indicates the presence of a hydrophobic amino
25 acid. The numerals located between amino acid residues of the motifs indicates the length of gaps in the sequences. For example, the "100" shown between motifs A and B reflects a 100 amino acid gap in the sequence between the motifs.

As noted above, Genbank searches identified a yeast protein (Genbank accession #u20618), and gene L8543.12 (Est2) containing or encoding amino acid
30 sequence that shows some homology to the *E. aediculatus* 123 kDa telomerase subunit. Based on the observations that both proteins contain reverse transcriptase

motifs in their C-terminal regions; both proteins share similarity in regions outside the reverse transcriptase motif; the proteins are similarly basic (pI = 10.1 for *E. aediculatus* and pI=10.0 for the yeast); and both proteins are large (123 kDa for *E. aediculatus* and 103 kDa for the yeast), these sequences comprise the catalytic core of their respective telomerases. It was contemplated based on this observation of homology in two phylogenetically distinct organisms as *E. aediculatus* and yeast, that human telomerase would contain a protein that has the same characteristics (i.e., reverse transcriptase motifs, is basic, and large [> 100 kDa]).

ii) Comparisons with the 43 kDa *E. aediculatus* Telomerase Subunit

The amino acid sequence of the "La-domain" of the 43 kDa *Euplotes aediculatus* polypeptide was compared with the sequence of the 95 kDa telomerase protein subunit of *Tetrahymena thermophila* (described above) to investigate their similarity. This sequence comparison is shown in Figure 38, while the *T.*

thermophila sequence is the lower sequence. The observed identity was determined to be approximately 23%, while the percent similarity was approximately 46%, values similar to what would be observed with any random protein sequence.

The amino acid sequence of the "La-domain" of the 43 kDa *Euplotes aediculatus* polypeptide was compared with the sequence of the 80 kDa telomerase protein subunit of *Tetrahymena thermophila* (described above) to investigate their similarity. This sequence comparison is shown in Figure 39. In this figure, the *E. aediculatus* sequence is the upper sequence, while the *T. thermophila* sequence is the lower sequence. The observed identity was determined to be approximately 26%, while the percent similarity was approximately 49%, values similar to what would be observed with any random protein sequence.

The amino acid sequence of a domain of the 43 kDa *E. aediculatus* polypeptide was also compared with La proteins from various other organisms. These comparisons are shown in Figure 41. In this Figure, highly conserved residues are indicated by white letters on a black background. Residues of the *E. aediculatus* sequences that are conserved in the other sequence are indicated in bold.

N. Identification of Telomerase Protein Subunits in Another Organism

In this Example, the sequences identified in the previous Examples above were used to identify the telomerase protein subunits of *Oxytricha trifallax*, a ciliate that is very distantly related to *E. aediculatus*. Primers were chosen based on the conserved region of the *E. aediculatus* 123 kDa polypeptide which comprised the reverse transcriptase domain motifs. Suitable primers were synthesized and used in a PCR reaction with total DNA from *Oxytricha*. The *Oxytricha* DNA was prepared according to methods known in the art. The PCR products were then cloned and sequenced using methods known in the art.

The oligonucleotide sequences used as the primers were as follows:
5'-(T/C)A(A/G)AC(T/A/C)AA(G/A)GG(T/A/C)AT(T/C)CC(C/T/A)(C/T)A(G/A)GG-3' and 5'-(G/A/T)GT(G/A/T)ATNA(G/A)NA(G/A)(G/A)TA(G/A)TC(G/A)TC-3'. Positions that were degenerate are shown in parentheses, with the alternative bases shown within the parenthesis. "N" represents any of the four nucleotides.

In the PCR reaction, a 50 µl reaction contained 0.2 mM dNTPs, 0.3 µg *Oxytricha trifallax* chromosomal DNA, 1 µl *Taq* polymerase (Boehringer-Mannheim), 2 micromolar of each primer, 1x reaction buffer (Boehringer-Mannheim). The reaction was incubated in a thermocycler (Perkin-Elmer) under the following conditions: 5 min at 95°C, 30 cycles consisting of 1 min at 94°C, 1 min at 53°C, and 1 min at 72°C, followed by a 10 min incubation at 72°C. The PCR-product was gel-purified and sequenced by the dideoxy-method (e.g., Sanger *et al.*, Proc. Natl. Acad. Sci. 74, 5463-5467 (1977)).

The deduced amino acid sequence of the PCR product was determined and compared with the *E. aediculatus* sequence. Figure 47 shows the alignment of these sequences, with the *O. trifallax* sequence shown in the top row, and the *E. aediculatus* sequence shown in the bottom row. As can be seen from this figure, there is a great deal of homology between the *O. trifallax* polypeptide sequence identified in this Example with the *E. aediculatus* polypeptide sequence. Thus, it is

clear that the sequences identified in the present invention are useful for the identification of homologous telomerase protein subunits in other eukaryotic organisms. Indeed, development of the present invention has identified homologous telomerase sequences in multiple, diverse species, as described herein.

5

O. Identification of *Tetrahymena* Telomerase Sequences

In this Example, a *Tetrahymena* clone was produced that shares homology with the *Euplotes* sequences, and EST2p.

This experiment utilized PCR with degenerate oligonucleotide primers directed against conserved motifs to identify regions of homology between *Tetrahymena*, *Euplotes*, and EST2p sequences. The PCR method used in this Example is a novel method designed to amplify specifically rare DNA sequences from complex mixtures. This method avoids the problem of amplification of DNA products with the same PCR primer at both ends (*i.e.*, single primer products) commonly encountered in PCR cloning methods. These single primer products produce unwanted background and can often obscure the amplification and detection of the desired two-primer product. The method used in this experiment preferentially selects for two-primer products. In particular, one primer is biotinylated and the other is not. After several rounds of PCR amplification, the products are purified using streptavidin magnetic beads and two primer products are specifically eluted using heat denaturation. This method finds use in settings other than the experiments described in this Example. Indeed, this method finds use in application in which it is desired to specifically amplify rare DNA sequences, including the preliminary steps in cloning methods such as 5' and 3' RACE, and any method that uses degenerate primers in PCR.

A first PCR run was conducted using *Tetrahymena* template macronuclear DNA isolated using methods known in the art, and the 24-mer forward primer with the sequence 5' biotin-GCCTATTT(TC)TT(TC)TA(TC)(GATC)(GATC)(GATC)AC(GATC)GA-3' designated as "K231," corresponding to the FFYXTE region, and the 23-mer reverse primer with the sequence 5'- CCAGATAT(GATC)A(TGA)(GATC)A(AG)(AG)AA(AG)TC(AG)TC-3', designated as "K220,"

corresponding to the DDFL(FIL)I region. This PCR reaction contained 2.5 µl DNA (50 ng), 4 µl of each primer (20 µM), 3 µl 10x PCR buffer, 3 µl 10x dNTPs, 2 µl Mg, 0.3 µl *Taq*, and 11.2 µl dH₂O. The mixture was cycled for 8 cycles of 94°C for 45 seconds, 37°C for 45 seconds, and 72 °C for 1 minute.

5 This PCR reaction was bound to 200 µl streptavidin magnetic beads, washed with 200 µl TE, resuspended in 20 µl dH₂O and then heat-denatured by boiling at 100°C for 2 minutes. The beads were pulled down and the eluate removed. Then, 2.5 µl of this eluate was subsequently reamplified using the above conditions, with the exception being that 0.3 µl of α -³²P dATP was included, and the
10 PCR was carried out for 33 cycles. This reaction was run a 5% denaturing polyacrylamide gel, and the appropriate region was cut out of the gel. These products were then reamplified for an additional 34 cycles, under the conditions listed above, with the exception being that a 42°C annealing temperature was used.

A second PCR run was conducted using *Tetrahymena* macronuclear
15 DNA template isolated using methods known in the art, and the 23-mer forward primer with the sequence 5'-

ACAATG(CA)G(GATC)(TCA)T(GATC)(TCA)T(GATC)CC
(GATC)AA(AG)AA-3' , designated as "K228," corresponding to the region
R(LI)(LI)PKK , and a reverse primer with the sequence 5'-

20 ACGAATC(GT)(GATC)GG
(TAG)AT(GATC)(GC)(TA)(AG)TC(AG)TA(AG)CA 3' , designated "K224,"
corresponding to the CYDSIPR region. This PCR reaction contained 2.5 µl DNA (50 ng), 4 µl of each primer (20 µM), 3 µl 10x PCR buffer, 3 µl 10x dNTPs, 2 µl Mg, 0.3 µl α -³²P dATP, 0.3 µl *Taq*, and 10.9 µl dH₂O. This reaction was run on a
25 5% denaturing polyacrylamide gel, and the appropriate region was cut out of the gel. These products were reamplified for an additional 34 cycles, under the conditions listed above, with the exception being that a 42°C annealing temperature was used.

Ten µl of the reaction product from run 1 were bound to streptavidin-coated magnetic beads in 200 µl TE. The beads were washed with 200 µl TE, and
30 then resuspended in 20 µl of dH₂O, heat denatured, and the eluate was removed. The

reaction product from run 2 was then added to the beads and diluted with 30 μ l 0.5x SSC. The mixture was heated from 94°C to 50°C. The eluate was removed and the beads were washed three times in 0.5x SSC at 55°C. The beads were then resuspended in 20 μ l dH₂O, heat denatured, and the eluate was removed, designated as "round 1 eluate" and saved.

To isolate the *Tetrahymena* band, the round 1 eluate was reamplified with the forward primer K228 and reverse primer K227 with the sequence 5'-CAATTCTC(AG)TA(AG)CA(GATC)(CG)(TA)(CT)TT(AGT)AT(GA)TC-3', corresponding to the DIKSCYD region. The PCR reactions were conducted as described above. The reaction products were run on a 5% polyacrylamide gel; the band corresponding to approximately 295 nucleotides was cut from the gel and sequenced.

The clone designated as 168-3 was sequenced. The DNA sequence (including the primer sequences) was found to be:

15 GATTACTCCCGAAGAAAGGATCTTTCCGTCCAATCATGACTTTCTTAAGAA
AGGACAAGCAAAAAAATATTAAGTTAAATCTAAATTAAATTCTAATGGATA
GCCAACTTGTGTTTAGGAATTTAAAAGACATGCTGGGATAAAAGATAGGAT
ACTCAGTCTTTGATAATAAACAAATTCAGAAAAATTTGCCTAATTCATAG
AGAAATGGAAAAATAAAGGAAGACCTCAGCTATATTATGTCACTCTAGAC
20 ATAAAGACTTGCTAC.

Additional sequence of this gene was obtained by PCR using one unique primer designed to match the sequence from 168-3 ("K297" with the sequence 5'-GAGTGACATAATATACGTGA-3'; and the K231 (FFYXTE) primer. The sequence of the fragment obtained from this reaction, together with 168-3 is as follows (without the primer sequences):

25 AAACACAAGGAAGGAAGTCAAATATTCTATTACCGTAAACCAATATGGAA
ATTAGTGAGTAAATTAAGTCAAGTAAGAATTTAGTTTTCTGAAAA
GAATAAATAAATGAAAAATAATTTTATCAAAAAATTTAGCTTGAAGAGGA
GAATTTGGAAAAAGTTGAAGAAAAATTGATACCAGAAGATTCATTTTAGAA
30 ATACCCTCAAGGAAAGCTAAGGATTATACCTAAAAAAGGATCTTTCCGTCC

AATCATGACTTTCTTAAGAAAGGACAAGCAAAAAAATATTAAGTTAAATCT
 AAATTAAATTCTAATGGATAGCCAACTTGTGTTTAGGAATTTAAAAGACAT
 GCTGGGATAAAAGATAGGATACTCAGTCTTTGATAATAAACAAATTTTCAGA
 AAAATTTGCCTAATTCATAGAGAAATGGAAAAATAAAGGAAGACCTCAGC
 5 TATATTATGTCACTCTA.

The amino acid sequence corresponding to this DNA fragment was found to be:

KHKEGSQIFYRKPIWKLVSCLTIVKVRIQFSEKNKQMKNNFYQKIQLEENLE
 KVEEKLIPEDSFQKYPQGKLRIIPKKGSRPIMTFLRKDKQKNIKLNLNQILMDS
 10 QLVFRNLKDMLGQKIGYSVFDNKQISEKFAQFIEKWKNKGRPQLYYVTL.

This amino acid sequence was then aligned with other telomerase genes (EST2p, and *Euplotes*). The alignment is shown in Figure 53. A consensus sequence is also shown in this Figure.

15 P. Identification of *Schizosaccharomyces pombe* Telomerase Sequences

In this Example, the *tez1* sequence of *S. pombe* was identified as a homolog of the *E. aedicularis* p123, and *S. cerevisiae* Est2p.

Figure 55 provides an overall summary of these experiments. In this Figure, the top portion (Panel A) shows the relationship of two overlapping genomic clones, and the 5825 bp portion that was sequenced. The region designated at
 20 "*tez1*⁺" is the protein coding region, with the flanking sequences indicated as well, the box underneath the 5825 bp region is an approximately 2 kb *Hind*III fragment that was used to make the *tez1* disruption construct, as described below.

The bottom half of Figure 55 (Panel B) is a "close-up" schematic of
 25 this same region of DNA. The sequence designated as "original PCR" is the original degenerate PCR fragment that was generated with a degenerate oligonucleotide primer pair designed based on *Euplotes* sequence motif 4 (B') and motif 5 (C), as described.

30

i) PCR With Degenerate Primers

PCR using degenerate primers was used to find the homolog of the *E.*

aediculatus p123 in *S. pombe*. Figure 56 shows the sequences of the degenerate primers (designated as "poly 4" and "poly 1") used in this reaction. The PCR runs were conducted using the same buffer as described in previous Examples (See e.g., Part K, above), with a 5 minute ramp time at 94°C, followed by 30 cycles of 94°C for 30 seconds, 50°C for 45 seconds, and 72°C for 30 seconds, and 7 minutes at 72°C, followed by storage at 4°C. PCR runs were conducted using varied conditions, (i.e., various concentrations of *S. pombe* DNA and MgCl₂ concentrations). The PCR products were run on agarose gels and stained with ethidium bromide as described above. Several PCR runs resulted in the production of three bands (designated as "T," "M," and "B"). These bands were re-amplified and run on gels using the same conditions as described above. Four bands were observed following this re-amplification ("T," "M1," "M2," and "B"), as shown in Figure 57. These four bands were then re-amplified using the same conditions as described above. The third band from the top of the lane in Figure 57 was identified as containing the correct sequence for a telomerase protein. The PCR product designated as M2 was found to show a reasonable match with other telomerase proteins, as indicated in Figure 58. In addition to the alignment shown, this Figure also shows the actual sequence of *tez1*. In this Figure, the asterisks indicate residues shared with all four sequences (*Oxytricha* "Ot"; *E. aediculatus* "Ea_p123"; *S. cerevisiae* "Sc_p103"; and M2), while the circles (i.e., dots) indicate similar amino acid residues.

ii) 3' RT PCR

To obtain additional sequence information, 3' and 5' RT PCR were conducted on the telomerase candidate identified in Figure 58. Figure 59 provides a schematic of the 3' RT PCR strategy used. First, cDNA was prepared from mRNA using the oligonucleotide primer "Q_T," (5'-CCA GTG AGC AGA GTG ACG AGG ACT CGA GCT CAA GCT TTT TTT TTT TTT TT-3'), then using this cDNA as a template for PCR with "Q_O" (5'-CCA GTG AGC AGA GTG ACG-3'), and a primer designed based on the original degenerated PCR reaction (i.e., "M2-T" with the sequence 5'-G TGT CAT TTC TAT ATG GAA GAT TTG ATT GAT-G-3'). The

second PCR reaction (*i.e.*, nested PCR) with "Q₁" (5'-GAG GAC TCG AGC TCA AGC-3'), and another PCR primer designed with sequence derived from the original degenerate PCR reaction or "M2-T2" (5'-AC CTA TCG TTT ACG AAA AAG AAA GGA TCA GTG-3'). The buffers used in this PCR were the same as described
5 above, with amplification conducted beginning with a ramp up of 94° for 5 min, followed by 30 cycles of 94° for 30 sec, 55°C for 30 sec, and 72°C for 3 min, followed by 7 minutes at 72°C. The reaction products were stored at 4°C until use.

iii) Screening of Genomic and cDNA Libraries

10 After obtaining this additional sequence information, several genomic and cDNA libraries were screened to identify any libraries that contain this telomerase candidate gene. The approach used, as well as the libraries and results are shown in Figure 60. In this Figure, Panel A lists the libraries tested in this experiment; Panel B shows the regions used; Panels C and D show the dot blot
15 hybridization results obtained with these libraries. Positive libraries were then screened by colony hybridization to obtain genomic and cDNA version of *tez1* gene. In this experiment, approximately 3×10^4 colonies from the *Hind*III genomic library were screened and six positive clones were identified (approximately 0.01%). DNA was then prepared from two independent clones (A5 and B2). Figure 61 shows the
20 results obtained with the *Hind*III-digested A5 and B2 positive genomic clones.

In addition, cDNA REP libraries were used. Approximately 3×10^5 colonies were screened, and 5 positive clones were identified (0.002%). DNA was prepared from three independent clones (2-3, 4-1, and 5-20). In later experiments, it was determined that clones 2-3 and 5-20 contained identical inserts.

25

iv) 5' RT PCR

As the cDNA version of gene produced to this point was not complete, 5' RT-PCR was conducted to obtain a full length clone. The strategy is schematically shown in Figure 62. In this experiment, cDNA was prepared using
30 DNA oligonucleotide primer "M2-B" (5'-CAC TGA TCC TTT CTT TTT CGT AAA CGA TAG GT-3') and "M2-B2" (5'-C ATC AAT CAA ATC TTC CAT ATA GAA

ATG ACA-3'), designed from known regions of *tez1* identified previously. An oligonucleotide linker PCR Adapt SfiI with a phosphorylated 5' end ("P") (P-GGG CCG TGT TGG CCT AGT TCT CTG CTC-3'; was then ligated at the 3' end of this cDNA, and this construct was used as the template for nested PCR. In the first round of PCR, PCR Adapt SFI and M2-B were used as the primers; while PCR Adapt SfiII (5-GAG GAG GAG AAG AGC AGA GAA CTA GGC CAA CAC GCC CC-3'), and M2-B2 were used as primers in the second round. Nested PCR was used to increase specificity of reaction.

10

v) Sequence Alignments

Once the sequence of *tez1* was identified, it was compared with sequences previously described. Figure 63 shows the alignment of RT domains from telomerase catalytic subunits of *S. pombe* ("S.p. Tez1p"), *S. cerevisiae* ("S.c. Est2p"), and *E. aediculatus* p123 ("E.a. p123"). In this Figure, "h" indicates hydrophobic residues, while "p" indicates small polar residues, and "c" indicates charged residues. The amino acid residues indicated above the alignment show a known consensus RT motif of Y. Xiong and T.H. Eickbush (Y. Xiong and T.H. Eickbush, EMBO J., 9: 3353-3362 [1990]). The asterisks indicate the residues that are conserved for all three proteins. "Motif O" is identified herein and in Figure 63 as a motif specific to this telomerase subunit and not found in reverse transcriptases in general. It is therefore valuable in identifying other amino acid sequences as telomerase catalytic subunits.

Figure 64 shows the alignment of entire sequences from *Euplotes* ("Ea_p123"), *S. cerevisiae* ("Sc_Est2p"), and *S. pombe* ("Sp_Tez1p"). In Panel A, the shaded areas indicate residues shared between two sequences. In Panel B, the shaded areas indicate residues shared between all three sequences.

25

vi) Genetic Disruption of *tez1*

In this Example, the effects of disruption of *tez1* were investigated.

As telomerase is involved in telomere maintenance, it was hypothesized that if *tez1*

were indeed a telomerase component, disruption of *tez1* would cause gradual telomere shortening.

In these experiments, homologous recombination was used to disrupt the *tez1* gene in *S. pombe* specifically. This approach is schematically illustrated in Figure 65. As indicated in Figure 65, wild type *tez1* was replaced with a fragment containing the *ura4* or *LEU2* marker.

The disruption of *tez1* gene was confirmed by PCR (Figure 66), and a Southern blot was performed to check for telomere length. Figure 67 shows the Southern blot results for this experiment. Because an *ApaI* restriction enzyme site is present immediately adjacent to telomeric sequence in *S. pombe*, *ApaI* digestion of *S. pombe* genomic DNA preparations permits analysis of telomere length. Thus, DNA from *S. pombe* was digested with *ApaI* and the digestion products were run on an agarose gel and probed with a telomeric sequence-specific probe to determine whether the telomeres of disrupted *S. pombe* cells were shortened. The results are shown in Figure 67. From these results, it was clear that disruption of the *tez1* gene caused a shortening of the telomeres.

Q. Cloning and Characterization of Human Telomerase Protein and cDNA

In this Example, the nucleic and amino acid sequence information for human telomerase was determined. Partial homologous sequences were first identified in a BLAST search conducted using the *Euplotes* 123 kDa peptide and nucleic acid sequences, as well as *Schizosaccharomyces* protein and corresponding cDNA (*tez1*) sequences. The human sequences (also referred to as "hTCP1.1") were identified from a partial cDNA clone (clone 712562). Sequences from this clone were aligned with the sequences determined as described in previous Examples.

Figure 1 shows the sequence alignment of the *Euplotes* ("p123"), *Schizosaccharomyces* ("*tez1*"), *Est2p* (i.e., the *S. cerevisiae* protein encoded by the *Est2* nucleic acid sequence, and also referred to herein as "L8543.12"), and the human homolog identified in this comparison search. Figure 51 shows the amino acid sequence of *tez1*, while Figure 52 shows the DNA sequence of *tez1*. In Figure 52, the introns and other non-coding regions, are shown in lower case, while the

exons (*i.e.*, coding regions) are shown in upper case.

As shown in the Figures, there are regions that are highly conserved among these proteins. For example, as shown in Figure 1, there are regions of identity in "Motif 0," "Motif 1," "Motif 2," and "Motif 3." The identical amino acids are indicated with an asterisk (*), while the similar amino acid residues are indicated by a circle (•). This indicates that there are regions within the telomerase motifs that are conserved among a wide variety of eukaryotes, ranging from yeast to ciliates to humans. It is contemplated that additional organisms will likewise contain such conserved regions of sequence. Figure 49 shows the partial amino acid sequence of the human telomerase motifs, while Figure 50 shows the corresponding DNA sequence.

Sanger dideoxy sequencing and other methods were used, as known in the art to obtain complete sequence information of clone 712562. Some of the primers used in the sequencing are shown in Table 7. These primers were designed to hybridize to the clone), based on sequence complementarity to either plasmid backbone sequence or the sequence of the human cDNA insert in the clone.

Table 7. Primers

Primer	Sequence
TCP1.1	GTGAAGGCACTG TTCAGCG
TCP1.2	GTGGATGATTTCTTGTTGG
TCP1.3	ATGCTCCTGCGTTTG GTG
TCP1.4	CTGGACACTCAGCCCTTGG
TCP1.5	GGCAGGTGTGCTGGACACT
TCP1.6	TTTGATGATGCTGGCGATG
TCP1.7	GGGGCTCGTCTTCTACAGG
TCP1.8	CAGCAGGAGGATCTTGTAG

TCP1.9	TGACCCCAGGAGTGGCACG
TCP1.10	TCAAGCTGACTCGACACCG
TCP1.11	CGGCGTGACAGGGCTGC
TCP1.12	GCTGAAGGCTGAGTGTCC
TCP1.13	TAGTCCATGTTCACAATCG

From these experiments, it was determined that the EcoRI-NotI insert of clone 712562 contains only a partial open reading frame for the human telomerase protein, although it may encode an active fragment of that protein. The open reading frame in the clone encodes an approximately 63 kD protein. The sequence of the longest open reading frame identified is shown in Figure 68. The ORF begins at the ATG codon with the "met" indicated in the Figure. The poly A tail at the 3' end of the sequence is also shown. Figure 69 shows a tentative, preliminary alignment of telomerase reverse transcriptase proteins from the human sequence (human Telomerase Core Protein 1, "Hs TCP1"), *E. aediculatus* p123 ("Ep p123"), *S. pombe* tez1 ("Sp Tez1"), *S. cerevisiae* EST2 (Sc Est2"), and consensus sequence. In this Figure various motifs are indicated.

To obtain a full-length clone, probing of a cDNA library and 5'-RACE were used to obtain clones encoding portions of the previously uncloned regions. In these experiments, RACE (Rapid Amplification of cDNA Ends; See e.g., M.A. Frohman, "RACE: Rapid Amplification of cDNA Ends," in Innis *et al.* (eds), *PCR Protocols: A Guide to Methods and Applications* [1990], pp. 28-38; and Frohman *et al.*, Proc. Natl. Acad. Sci., 85:8998-9002 [1988]) was used to generate material for sequence analysis. Four such clones were generated and used to provide additional 5' sequence information (pFWRP5, 6, 19, and 20).

In addition, human cDNA libraries (inserted into lambda) were probed with the EcoRI-NotI fragment of the clone. One lambda clone, designated "lambda 25-1.1" (ATCC accession #209024), was identified as containing complementary

sequences. Figure 75 shows a restriction map of this lambda clone. The human cDNA insert from this clone was subcloned as an *EcoRI* restriction fragment into the *EcoRI* site of commercially available phagemid pBluescriptIIISK+ (Stratagene), to create the plasmid "pGRN121," which was deposited with the ATCC (ATCC
5 accession #209016). Preliminary results indicated that plasmid pGRN121 contains the entire open reading frame (ORF) sequence encoding the human telomerase protein.

The cDNA insert of plasmid pGRN121 was sequenced using techniques known in the art. Figure 70 provides a restriction site and function map
10 of plasmid pGRN121 identified based on this preliminary work. The results of this preliminary sequence analysis are shown in Figure 71. From this analysis, and as shown in Figure 70, a putative start site for the coding region was identified at approximately 50 nucleotides from the *EcoRI* site (located at position 707), and the location of the telomerase-specific motifs, "FFYVTE" , "PKP," "AYD," "QG", and
15 "DD," were identified, in addition to a putative stop site at nucleotide #3571 (See, Figure 72, which shows the DNA and corresponding amino acid sequences for the open reading frames in the sequence ("a", "b", and "c"). However, due to the preliminary nature of the early sequencing work, the reading frames for the various motifs were found not to be in alignment.

20 Additional analysis conducted on the pGRN121 indicated that the plasmid contained significant portions from the 5'-end of the coding sequence not present on clone 712562. Furthermore, pGRN121 was found to contain a variant coding sequence that includes an insert of approximately 182 nucleotides. This insert was found to be absent from the clone. As with the *E. aediculatus* sequences, such
25 variants can be tested in functional assays, such as telomerase assays to detect the presence of functional telomerase in a sample.

Further sequence analysis resolved the cDNA sequence of pGRN121 to provide a contiguous open reading frame that encodes a protein of molecular weight of approximately 127,000 daltons, and 1132 amino acids as shown in Figure
30 74. A refined map of pGRN121 based on this analysis, is provided in Figure 73. The results of additional sequence analysis of the hTERT cDNA are presented in

Figure 16 (SEQUENCE ID NO: 1).

5

EXAMPLE 2 (Technical Guidance)

CORRELATION OF hTERT ABUNDANCE AND CELL IMMORTALITY

The relative abundance of hTERT mRNA was assessed in six telomerase-negative mortal cell strains and six telomerase-positive immortal cell lines (Figure 5).

10 The steady state level of hTERT mRNA was significantly increased in immortal cell lines that had previously been shown to have active telomerase. Lower levels of the hTERT mRNA were detected in some telomerase-negative cell strains.

RT-PCR for hTERT, hTR, TP1 (telomerase-associated protein related to *Tetrahymena* p80 [Harrington et al., 1997, *Science* 275:973; Nakayama et al., 1997, *Cell* 88:875]) and GAPDH (to normalize for equal amounts of RNA template) was
15 carried out on RNA derived from the following cells: (1) human fetal lung fibroblasts GFL, (2) human fetal skin fibroblasts GFS, (3) adult prostate stromal fibroblasts 31 YO, (4) human fetal knee synovial fibroblasts HSF, (5) neonatal foreskin fibroblasts BJ, (6) human fetal lung fibroblasts IMR90, and immortalized cell lines: (7) melanoma LOX IMVI, (8) leukemia U251, (9) NCI H23 lung carcinoma, (10) colon adenocarcinoma
20 SW620, (11) breast tumor MCF7, (12) 293 adenovirus E1 transformed human embryonic kidney cell line.

hTERT nucleic acid was amplified from cDNA using oligonucleotide primers LT5 and LT6 (Table 2) for a total of 31 cycles (94°C 45s, 60°C 45s, 72°C 90s).

GAPDH was amplified using primers KI36 (5'-CTCAGACACCATGGGGAA
25 GGTGA) and K137 (5'-ATGATCTTGAGGCTGTTGTCATA) for a total of 16 cycles (94°C 45 s, 55°C 45 s, 72°C 90 s). hTR was amplified using primers F3b (5'-TCTAA CCCTAACTGAGAAGGGCGTAG) and R3c (5'-GTTTGCTCTAGAATGAACGGTG
GAAG) for a total of 22 cycles (94 °C 45s, 55 °C 45 s, 72 °C 90s). TP1 mRNA was amplified using primers TP1.1 and TP1.2 for 28 cycles (cycles the same as hTERT).

30 Reaction products were resolved on an 8% polyacrylamide gel, stained with SYBR Green (Molecular Probes) and visualized by scanning on a Storm 860 (Molecular

Dynamics). The results, shown in Figure 5, demonstrate that hTRT mRNA levels correlate directly with telomerase activity levels in the cells tested.

5

EXAMPLE 3 (Technical Guidance)

CHARACTERIZATION OF AN hTRT INTRONIC SEQUENCE

A putative intron was first identified by PCR amplification of human genomic DNA, as described in this example, and subsequently confirmed by sequencing the genomic clone λ G ϕ 5 (see Example 4). PCR amplification was carried out using the forward primer TCP1.57 paired individually with the reverse primers TCP1.46, TCP1.48, TCP1.50, TCP1.52, TCP1.54, TCP1.56, and TCP1.58 (see Table 2). The products from genomic DNA of the TCP1.57/TCP1.46, TCP1.48, TCP1.50, TCP1.52, TCP1.54, or TCP1.56 amplifications were approximately 100 basepairs larger than the products of the pGRN121 amplifications. The TCP1.57/TCP1.58 amplification was the same on either genomic or pGRN121 DNA. This indicated the genomic DNA contained an insertion between the sites for TCP1.58 and TCP1.50. The PCR products of TCP1.57/TCP1.50 and TCP1.57/TCP1.52 were sequenced directly, without subcloning, using the primers TCP1.39, TCP1.57, and TCP1.49.

20 As shown below, the 104-base intronic sequence (SEQUENCE ID NO: 7) is inserted in the hTRT mRNA (shown in bold) at the junction corresponding to bases 274 and 275 of Figure 16:

CCCCCGCCGCCCCCTCCTTCCGCCAG/GTGGGCCTCCCCGGGGTCGGCG
TCCGGCTGGGGTTGAGGGCGGCCGGGGGAACCAGCGACATGCGGAGAGC
25 AGCGCAGGCGACTCAGGGCGCTTCCCCCGCAG/GTGTCTCCTGCCTGAAGGA
GCTGGTGGCCCGAGTGCTGCAG

The "/" indicates the splice junctions; the sequence shows good matches to consensus 5' and 3' splice site sequences typical for human introns.

30 This intron contains motifs characteristic of a topoisomerase II cleavage site and a NF κ B binding site (see Figure 21). These motifs are of interest, in part,

because expression of topoisomerase II is up regulated in most tumors. It functions to relax DNA by cutting and rewinding the DNA, thus increasing expression of particular genes. Inhibitors of topoisomerase II have been shown to work as anti-tumor agents. In the case of NF κ B, this transcription factor may play a role in regulation of telomerase during terminal differentiation, such as in early repression of telomerase during development and so is another target for therapeutic intervention to regulate telomerase activity in cells.

EXAMPLE 4

10 CLONING OF LAMBDA PHAGE G Φ 5 AND CHARACTERIZATION OF hTERT GENOMIC SEQUENCES

A. Lambda G Φ 5

A human genomic DNA library was screened by PCR and hybridization to identify a genomic clone containing hTERT RNA coding sequences. The library was a human fibroblast genomic library made using DNA from WI38 lung fibroblast cells (Stratagene, Cat # 946204). In this library, partial Sau3AI fragments are ligated into the XhoI site of Lambda FIX Φ II Vector (Stratagene), with an insert size of 9-22 kb.

The genomic library was divided into pools of 150,000 phage each, and each pool screened by nested PCR (outer primer pair TCP1.52 & TCP1.57; inner pair TCP1.49 & TCP1.50, see Table 1). These primer pairs span a putative intron (see Example 3, *supra*) in the genomic DNA of hTERT and ensured the PCR product was derived from a genomic source and not from contamination by the hTERT cDNA clone. Positive pools were further subdivided until a pool of 2000 phage was obtained. This pool was plated at low density and screened via hybridization with a DNA fragment encompassing basepairs 1552-2108 of Figure 16 (restriction sites SphI and EcoRV, respectively).

Two positive clones were isolated and rescreened via nested PCR as described above; both clones were positive by PCR. One of the clones (λ G Φ 5) was digested with NotI, revealing an insert size of approximately 20 kb. Subsequent mapping (see below) indicated the insert size was 15 kb and that phage G Φ 5 contains

approximately 13 kb of DNA upstream from the start site of the cDNA sequence.

Phage GΦ5 was mapped by restriction enzyme digestion and DNA sequencing. The resulting map is shown in Figure 7. The phage DNA was digested with *NcoI* and the fragments cloned into pBBS167. The resulting subclones were
5 screened by PCR to identify those containing sequence corresponding to the 5' region of the hTRT cDNA. A subclone (pGRN140) containing a 9 kb *NcoI* fragment (with hTRT gene sequence and 4-5 kb of lambda vector sequence) was partially sequenced to determine the orientation of the insert. pGRN 140 was digested using *SaI* to remove lambda vector sequences, resulting in pGRN144. pGRN144 was then sequenced. The
10 results of the sequencing are provided in Figure 21. The 5' end of the hTRT mRNA corresponds to base 2441 of Figure 21. As indicated in Figure 7, two Alu sequence elements are located 1700 base pairs upstream of the hTRT cDNA 5' end and provide a likely upstream limit to the promoter region of hTRT. The sequence also reveals an intron positioned at bases 4173 in Figure 21, 3' to the intron described in Example 3,
15 *supra*.

B. Additional Genomic Clones

In addition to the genomic clone described above, two P1 bacteriophage clones and one human BAC clone are provided as illustrative embodiments of the
20 invention. P1 inserts are usually 75-100 kb, and BAC inserts are usually over 100 Kb.

The P1 clones (DMPC-HFF#1-477(F6) -GS #15371 and DMPC-HEF#1-1103(H6) -GS #15372) were obtained by PCR screening of a human P1 library derived from human foreskin fibroblast cells (Shepherd et al., 1994, *PNAS USA* 91:2629) using primers TCP1.12 and UTR2 which amplify the 3' end of hTRT. These
25 clones were both negative (failed to amplify) with primers that amplify the 5' end of hTRT.

The human BAC clone (326 E 20) was obtained with a hybridization screen of a BAC human genomic library using an 1143 bp *SphI/XmnI* fragment of pGRN121 (Figure 16; bases 1552-2695) that encompasses the RT motif region. The
30 clone is believed to include the 5' end of the gene. The hTRT genomic clones in this example are believed to encompass the entire hTRT gene.

5

EXAMPLE 5 (Technical Guidance)

CHROMOSOMAL LOCATION OF hTRT GENE

The hTRT gene was localized to chromosome 5p by radiation hybrid mapping (Boehnke et al., 1991, *Am J Hum Genet* 49:1174; Walter et al., 1994, *Nature Genet* 7:22) using the medium resolution Stanford G3 panel of 83 RH clones of the whole human genome (created at the Stanford Human Genome Center). A human lymphoblastoid cell line (donor; rM) was exposed to 10,000 rad of x-rays and was then fused with nonirradiated hamster recipient cells (A3). Eighty-three independent somatic cell hybrid clones were isolated, and each represents a fusion event between an irradiated donor cell and a recipient hamster cell. The panel of G3 DNA was used for ordering markers in the region of interest as well as establishing the distance between these markers.

The primers used for the RH mapping were TCP1.12 and UTR2 with amplification conditions of 94°C 45 sec, 55°C 45 sec, 72°C 45 sec, for 45 cycles using Boehringer Mannheim (RTM) Taq buffer and Perkin-Elmer Taq. The 83 pools were amplified independently and 14 (17%) scored positive for hTRT (by appearance of a 346 bp band). The amplification results were submitted to Stanford RH server, which then provided the map location, 5p, and the closest marker, STS D5S678.

By querying the Genethon genome mapping web site, the map location identified a YAC that contains the STS marker D5S678: CEPH YAC 780_C_3 Size: 390,660 kb. This YAC also contained chromosome 17 markers. This result indicated that the hTRT gene is on chromosome 5, near the telomeric end. There are increased copy numbers of 5p in a number of tumors. Cri-du-chat syndrome also has been mapped to deletions in this region.

30

EXAMPLE 6 (Technical Guidance except where indicated)

DESIGN AND CONSTRUCTION OF VECTORS FOR EXPRESSION OF hTRT

PROTEINS AND POLYNUCLEOTIDES

Expression of hTERT in Bacteria

The following portion of this example details the design of hTERT-
5 expressing bacterial and eukaryotic cell expression vectors to produce large quantities of
full-length, biologically active hTERT. Generation of biologically active hTERT protein in
this manner is useful for telomerase reconstitution assays, assaying for telomerase
activity modulators, analysis of the activity of newly isolated species of hTERT,
identifying and isolating compounds which specifically associate with hTERT, analysis of
10 the activity of an hTERT variant protein that has been site-specifically mutated, and as an
immunogen, as a few examples.

pThioHis A/hTERT Bacterial Expression Vector

To produce large quantities of full-length hTERT, the bacterial expression
15 vector pThioHis A (Invitrogen, San Diego, CA) was selected as an expression system.
The hTERT-coding insert includes nucleotides 707 to 4776 of the hTERT insert in the
plasmid pGRN121. This nucleotide sequence includes the complete coding sequence
for the hTERT protein.

This expression vector of the invention is designed for inducible
20 expression in bacteria. The vector can be induced to express, in *E. coli*, high levels of a
fusion protein composed of a cleavable, HIS tagged thioredoxin moiety and the full
length hTERT protein. The use of the expression system was in substantial accordance
with the manufacturer's instructions. The amino acid sequence of the fusion protein
encoded by the resulting vector of the invention is shown below; (-*-) denotes an
25 enterokinase cleavage site:

MSDKIIHLTDDSFDTDLKADGAILVDFWAHWCGPCKMIAPILDEIADEYQGKLTVAK
LRIDHNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSGSGDD
DDK--*-VPMHELEIFEFAAASTQRCVLLRTWEALAPATPAMPRAPCRAVRSLLRSHY
REVLPLATFVRRLGPQGWRVLVQRGDPAAFRALVAQCLVCVPWDARPPPAAPSFRQVSC
30 LKELVARVLQRLCERGAKNVLAFGFALLDGARGGPPEAFTTSVRSYLPNTVTDALRGS
GAWGLLLRRVGDDVLVHLLARCALFVLVAPSCAYQVCGPPLYQLGAATQARPPPHASG
PRRRLGCERAWNHSVREAGVPLGLPAPGARRRGSASRSLPLPKRPRRGAAPEPERTP

VGQGSWAHPGRTRGPSDRGFCVVSPARPAEEATSLEGALSGTRHSHPSVGRQHHAGPP
 STSRPPRPWDTPCPPVYAETKHFLYSSGDKEQLRPSFLLSSLRPSLTGARRLVETIFL
 GSRPWMPGTPRRLPRLPQRYWQMRPLFLELLGNHAQCPYGVLLKTHCPLRAAVTPAAG
 VCAREKPQGSVAAPEEEDTDPRRLVQLLRQHSSPWQVYGFVRACLRRLVPPGLWGSRH
 5 NERRFLRNTKKFISLGKHAKLSLQELTWKMSVRDCAWLRRSPGVGCVPAAEHRLREEI
 LAKFLHWLMSVYVVELLRSFFYVTETTFQKNRLFFYRKS VWSKLQSIGIRQH LKRVQL
 RELSEAEVRQHREARPALLTSRLRFIPKPDGLRPIVNDYVVGARTFRREKRAERLTS
 RVKALFSVLNYERARRPGLLGASVLGLDDIHRWRTFVLRVRAQDPPPELYFVKVDVT
 GAYDTIPQDRLTEVIASIIKPQNTYCVRRYAVVQKAAHGHVRKAFKSHVSTLTDLQPY
 10 MRQFVAHLQETSPLRDAVVEQSSSLNEASSGLFDVFLRFMCHHAVRIRGKSYVQCQG
 IPQGSILSTLLCSLCYGD MENKLFAGIRRDGLLLRLVDDFLLVTPHLTHAKTFLRTL
 RGVPEYGCVVNLRKTVNFPVEDEALGGTAFVQMPAHGLFPWCGLLLDTRTLEVQSDY
 SSYARTSIRASLTFNRGFKAGRNMRRKLFGLVRLKCHSLFLDLQVNSLQTVCTNIYKI
 LLLQAYRFHACVLQLPFHQVWKNPTFFLRVISDTASLCYSILKAKNAGMSLGAKGAA
 15 GPLPSEAVQWLCHQAFLLKLTRHRVTYVPLLGSRLTAQTQLSRKLP GTTTLTALEAAAN
 PALPSDFKTILD

pGEX-2TK with hTRT Nucleotides 3272 to 4177 of pGRN121

This construct of the invention is used to produce fusion protein for,
 20 e.g., the purpose of raising polyclonal and monoclonal antibodies to hTRT protein.
 Fragments of hTRT can also be used for other purposes, such as to modulate
 telomerase activity, for example, as a dominant-negative mutant or to prevent the
 association of a telomerase component with other proteins or nucleic acids.

To produce large quantities of an hTRT protein fragment, the *E. coli*
 25 expression vector pGEX-2TK (Pharmacia Biotech, Piscataway N.J) was selected, and
 used essentially according to manufacturer's instructions to make an expression
 vector of the invention. The resulting construct contains an insert derived from
 nucleotides 3272 to 4177 of the hTRT insert in the plasmid pGRN121. The vector
 directs expression in *E. coli* of high levels of a fusion protein composed of
 30 glutathione-S-transferase sequence (underlined below), thrombin cleavage sequence
 (double underlined), recognition sequence for heart muscle protein kinase (italicized),
 residues introduced by cloning in brackets ([GSVTK]) and hTRT protein fragment (in
 bold) as shown below:

MSPIILGYWKIKGLVQPTRLLLEYLEEKVEEHLVERDEGDKWRNKKFELGLEFPNLPYY

5 IDGDVKILTQSMALIRVIADKHNMLGGCPKERAETSMLEGAVLDIRYGVSRITAYSKDFE
TLKVDLSEKLPKMEEDRLCHKTYINGDHVTHPDLMYDALDVVLYMDPMCLDAEP
KLVCFKKRIEAIPOIDKYLKSSKYIAWPLOGWOATEGGGDHPPKSDLVPRGSRASV [
 10 GSVTK] IPQGSILSTLLCSLCYGD MENKLFAGIRRDGLLLRLVDDFLLVTPHLTHAKT
FLRTLVRGVPEYGCVVNL RKT VVNF PVEDEALGGTAFVQMPAHGLFPWCGLLLDTRTL
EVQSDYSSYARTSIRASVT FNRGFKAGRNMRRKLFGLRLKCHSLFLDLQVNSLQTV
TNIYKILLQAYRFHACVLQLPFHQQVWKNPTFFLRVISDTASLCYSILKAKNAGMSL
GAKGAAGPLPSEAVQWLCHQAFLLKLTRHRVTYVPLLGLSLRTAQTQLSRKLPGTTLTA
 15 LEAANPALPSDFKTILD

10 When this fusion protein was expressed, it formed insoluble
 aggregates. It was treated generally as described above, in the section entitled
 purification of proteins from inclusion bodies. Specifically, induced cells were
 suspended in PBS (20 mM sodium phosphate, pH 7.4, 150 mM NaCl) and disrupted
 15 by sonication. NP-40 was added to 0.1%, and the mixture was incubated for 30
 minutes at 4°C with gentle mixing. The insoluble material was collected by
 centrifugation at 25,000g for 30 minutes at 4°C. The insoluble material was washed
 once in 4M urea in PBS, collected by centrifugation, then washed again in PBS. The
 collected pellet was estimated to contain greater than 75% fusion protein. This
 20 material was dried in a speed vacuum, then suspended in adjuvant for injection into
 mice and rabbits for the generation of antibodies. Separation of the recombinant
 protein from the glutathione S-transferase moiety is accomplished by site-specific
 proteolysis using thrombin according to manufacturer's instructions.

25 *pGEX-2TK with hTRT Nucleotides 2426 to 3274 of pGRN121 with HIS-8 Tag*

To produce large quantities of a fragment of hTRT, another *E. coli*
 expression vector pGEX-2TK construct was prepared. This construct contains an
 insert derived from nucleotides 2426 to 3274 of the hTRT insert in the plasmid
 pGRN121 and a sequence encoding eight consecutive histidine residues (HIS-8 Tag).
 30 To insert the HIS-8 TAG, the pGEX-2TK vector with hTRT nucleotides 2426 to
 3274 of pGRN121 was linearized with BamH1. This opened the plasmid at the
 junction between the GST-thrombin-heart muscle protein kinase and the hTRT coding
 sequence. A double stranded oligonucleotide with BamH1 compatible ends was
 ligated to the linearized plasmid resulting in the in-frame introduction of eight
 35 histidine residues upstream of the hTRT sequence.

The vector directs expression in *E coli* of high levels of a fusion protein composed of glutathione-S-transferase sequence (underlined); thrombin cleavage sequence (double underlined); recognition sequence for heart muscle protein kinase (italicized); a set of three and a set of five residues introduced by cloning are in brackets ([GSV] and [GSVTK]); eight consecutive histidines (also double underlined); and hTRT protein fragment (in bold):

5
10
15
MSPILGYWKIKGLVOPTRLLEYLEEKYEEHL~~YERDEGDKWRNKKFELGLEFP~~
NLPYYIDGDVKLTQSM~~AIIRYIADKH~~NMLGGCPKERA~~EISM~~LEGAVLDIRYGVS
RIAYSKDFETLKVD~~FLSKLPEMLKMFEDRLCHKTYLNGDHVTHPDFM~~LYDALD
VVL~~YMDPMCLDAFPKLVC~~FKKRIEAIPOIDKYLKSSKYIAWPLOGWOATFGGG
DHPPKSDLVPRG~~SRRASV~~[GSV]HHHHHHHH[GSVTK]MSVYVVELLRSFFYVT
ETTFQKNRLFFYRPSVWSKLQSIGIRQHLKRVQLRELSEAEVRQHREARPA
LLTSRLRFIPKPDGLRP~~IVNMDYVVGARTFRREKRAERLTSRVKALF~~SVLN
YERARRPGLLGASVLGLDDIHRAWRTFVL~~RVRAQDPPPELYFVKVD~~VTGA
YDTIPQDRLTEVIASIIKPQNTYCVRRYA~~VVQKAAHGHVRKA~~FKSHVSTLT
DLQPYMRQFVAHLQETSPLRDAVVIEQSSSLNEASSGLFDVFLRFMCHHAV
RIRGKSYVQCQGI

Each of the pGEX-2TK vectors of the invention can be used to produce fusion protein for the purpose of raising polyclonal and monoclonal antibodies to hTRT protein. Additionally, this fusion protein can be used to affinity purify antibodies raised to hTRT peptides that are encompassed within the fusion protein. Separation of the recombinant protein from the glutathione S-transferase moiety can be accomplished by site-specific proteolysis using thrombin according to manufacturer's instructions.

25

pGEX-2TK with hTRT Nucleotides 2426 to 3274 of pGRN121, no HIS-8 Tag

To produce large quantities of a fragment of hTRT, another *E. coli* expression vector pGEX-2TK construct was prepared.

This construct contains an insert derived from nucleotides 2426 to 3274 of the hTRT insert in the plasmid pGRN121, but without the HIS-8 tag of the construct described above. The vector directs expression in *E coli* of high levels of a fusion protein composed of glutathione-S-transferase (underlined), thrombin cleavage sequence (double underlined), recognition sequence for heart muscle protein kinase (italicized), residues introduced by cloning in brackets ([GSVTK]) and hTRT protein fragment (in bold):

30
35

MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFPNLPY
 IDGDVKLTQSMALIRYIADKHNMLGGCPKERAELSMLEGAVLDIRYGVSRAYSKEFE
 TLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTHPDFMLYDALDVVLYMDPMCLDAEP
 5 KLVCFFKKRIEAIPOIDKYLKSSKYIAWPLQGWQATFEGGGDHPPKSDLVPRGSRRASV [
 GSVTK]MSVYVVELLRSFFYVTETTFQKNRLFFYRPSVWSKLQSIGIRQHLKRVQLRE
 LSEAEVRQHRARPALLTSRLRFIPKPDGLRPIVNMDYVVGARTFRREKRAERLTSRK
 ALFSVLNYERARRPGLLGASVLGLDDIHRAWRTFVLRVRAQDPPPEYFVKVDVTGAYD
 TIPQDRLTEVIASIIKPQNTYCVRRYAVVQKAHGVKAFKSHVSTLTDLQPYMRQFV
 10 AHLQETSPLRDAVVEQSSSLNEASGLFDVFLRFMCHHAVRIRGKSYVQCQGI

pGEX-2TK with hTRT Nucleotides 1625 to 2458 of pGRN121

To produce large quantities of a fragment of hTRT protein, another *E. coli* expression vector pGEX-2TK construct was prepared.

15 This construct contains an insert derived from nucleotides 1625 to
 2458 of the hTRT insert in the plasmid pGRN121. The vector directs expression in
E. coli of high levels of a fusion protein composed of glutathione-S-transferase,
 (underlined), thrombin cleavage sequence (double underlined), recognition sequence
 for heart muscle protein kinase (italicized) residues introduced by cloning in brackets
 20 ([GSVTK]) and hTRT protein fragment (in bold):

MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFP
 NLPYYIDGDVKLTQSMALIRYIADKHNMLGGCPKERAELSMLEGAVLDIRYGV
 SRIAYSKDFETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTHPDFMLYDA
 LDVVLYMDPMCLDAFPKLVCFFKKRIEAIPOIDKYLKSSKYIAWPLQGWQATFG
 25 GGDHPPKSDLVPRGSRRASV[GSVTK]ATSLEGALSGTRHSHPSVGRQHHAGP
 PSTSRPPRPWDTPCPPVYAETKHFLYSSGDKEQLRPSFLLSSLRPSLTGARR
 LVETIFLGSRPWMPGTPRRLPRLPQRYWQMRPLFLELLGNHAQCPYGVLL
 KTHCPLRAAVTPAAGVCAREKPQGSVAAPEEEDTDPRLVQLLRQHSSPW
 QVYGFVRACLRLVPPGLWGSRHNERFLRNTKKFISLGKHAKLSQLQELT
 30 WKMSVRDCAWLRRSPGVGCVPAAEHLRLREEILAKFLHWLMSVYVVELLR
 S

pGEX-2TK with hTRT Nucleotides 782 to 1636 of pGRN121

35 To produce large quantities of a fragment of hTRT protein, another *E. coli*
 expression vector pGEX-2TK construct was prepared.

This construct contains an insert derived from nucleotides 782 to 1636
 of the hTRT insert in the plasmid pGRN121. The vector directs expression in *E. coli*
 of high levels of a fusion protein composed of glutathione-S-transferase, (underlined),
 thrombin cleavage sequence (double underlined), recognition sequence for heart

muscle protein kinase (italicized) residues introduced by cloning in brackets

([GSVTK]) and hTRT protein fragment (in bold):

5 MSPTILGYWKIKGLVOPTRLLILEYLEEKVEEHLVERDEGDKWRNKKFELGLEFPNLPVY
IDGDVKLTQSMATIRYIADKHNMLGGCPKERAELSMLEGAVLDTRYGVSRIAYSKDFE
TLKVDFELSKLPEMLKMFEDRLCHKTYLNGDHVTHPDFMLYDALDVVLVMDPMCLDAFP
10 KLVCFKKRIFEAIPQIDKYLKSSKYIAWPIQGWQATFGGGDHPPKSDLVPRGSRASV
[GSVTK]MPRAPRCRAVRSLLSHYREVLPLATFVRRLGPGQWRLVQRGDPAAFRALVAQ
CLVCVPWDARPPAAPSFQVSCLELVARVLQRLCERGAKNVLAFGFALLDGARGGPP
EATTSVRSYLPNTVTDALRGSGAWGLLLRRVGDDVLVHLLARCALFVLVAPCAYQVCG
15 PPLYQLGAATQARPPPHASGPRRRLGCERAWNHSVREAGVPLGLPAPGARRRGGSASR
SLPLPKRPRRGAAPEPERTPVGQGSWAHPGTRGPSDRGFCVVSAPARPAEEATSL

pT7FLhTRT with hTRT cDNA Lacking 5'-Non-Coding Sequence

As described above, in one embodiment, the invention provides for an
15 hTRT that is modified in a site-specific manner to facilitate cloning into bacterial,
mammalian, yeast and insect expression vectors without any 5' untranslated hTRT
sequence. In some circumstances, minimizing the amount of non-protein encoding
sequence allows for improved protein production (yield) and increased mRNA
stability. In this embodiment of the invention, the hTRT gene's 5' non-coding region
20 was removed before cloning into a bacterial expression vector.

This was effected by engineering an additional restriction endonuclease
site just upstream (5') to the start (ATG) codon of the hTRT coding sequence
(Figure 16). The creation of a restriction site just 5' to the coding region of the
protein allows for efficient production of a wide variety of vectors that encode fusion
25 proteins, such as fusion proteins comprising labels and peptide TAGs, for
immunodetection and purification.

Specifically, the oligonucleotide

5'- CCGGCCACCCCCCATATGCCGCGCGCTCCC-3' was used as described above
to modify hTRT cDNA nucleotides 779 to 781 of the hTRT cDNA (Figure 16) from
30 GCG to CAT. These 3 nucleotides are the last nucleotides before the ATG start
codon so they do not modify the protein sequence. The change in sequence results in
the creation of a unique NdeI restriction site in the hTRT cDNA. Single-stranded
hTRT DNA was used as a DNA source for the site directed mutagenesis. The
resulting plasmid was sequenced to confirm the success of the mutagenesis.

35 This modification allowed the construction of the following plasmid of

the invention, designated pT7FLhTRT. The site-specifically modified hTRT sequence (addition of the NdeI restriction site) was digested with NdeI and NotI (and filled in with Klenow enzyme to generate blunt ended DNA) to generate an hTRT encoding nucleic acid fragment. The fragment was then cloned into a pSL3418
5 plasmid previously restriction digested with NdeI and SmaI (also a blunt ended cutter). pSL 3418 is a modified pAED4 plasmid into which a FLAG sequence (Immunex Corp, Seattle WA) and an enterokinase sequence are inserted just upstream from the above-referenced NdeI site. This plasmid, designated pT7FLhTR, allows the expression of full length hTRT (with a Flag-Tag at its 5' end) in an *E.coli* strain
10 expressing the T7 RNA polymerase.

Plasmids with hTRT cDNA Lacking 3'-Non-Coding Sequence

As discussed above, the invention provides for expression vectors containing TRT-encoding nucleic acids in which some or all non-coding sequences
15 have been deleted. In some circumstances, minimizing the amount of non-protein encoding sequence allows for improved protein production (yield) and increases mRNA stability. In this embodiment of the invention, the 3' untranslated region of hTRT is deleted before cloning into a bacterial expression plasmid.

The plasmid pGRN121, containing the full length hTRT cDNA, as
20 discussed above, was first deleted of all ApaI sites. This was followed by deletion of the MscI-HincII hTRT restriction digest enzyme fragment containing the 3'UTR. The NcoI-XbaI restriction digest fragment containing the stop codon of hTRT was then inserted into the NcoI-XbaI site of pGRN121 to make a plasmid equivalent to pGRN121, designated pGRN124, except lacking the 3'UTR.

25

Bacterial Expression Vectors Using Antibiotic Selection Markers

The invention also provides for bacterial expression vectors that can contain selection markers to confer a selectable phenotype on transformed cells and sequences coding for episomal maintenance and replication such that integration into
30 the host genome is not required. For example, the marker may encode antibiotic resistance, particularly resistance to chloramphenicol (see Harrod (1997) *Nucleic*

Acids Res. 25: 1720-1726), kanamycin, G418, bleomycin and hygromycin, to permit selection of those cells transformed with the desired DNA sequences, see for example, Blondelet-Rouault (1997) *Gene* 190:315-317; and Mahan (1995) *Proc Natl Acad Sci U S A* 92:669-673.

In one embodiment of the invention, the full length hTRT was cloned into a modified BlueScript plasmid vector (Stratagene, San Diego, CA), designated pBBS235, into which a chloramphenicol antibiotic resistance gene had been inserted. The NotI fragment from pGRN124 (discussed above) containing the hTRT ORF into the NotI site of pBBS235 so that the TRT ORF is in the opposite orientation of the vector's Lac promoter. This makes a plasmid that is suitable for mutagenesis of plasmid inserts, such as TRT nucleic acids of the invention. This plasmid construct, designated pGRN125, can be used in the methods of the invention involving mutagenesis of telomerase enzyme and TRT protein coding sequences and for *in vitro* transcription of hTRT using the T7 promoter (and *in vitro* transcription of antisense hTRT using the T3 promoter).

In another embodiment of the invention, NotI restriction digest fragments from pGRN124 containing the hTRT ORF were subcloned into the NotI site of pBBS235 (described above) so the TRT ORF is in the same orientation as the vector's Lac promoter. This makes a plasmid, designated pGRN126, that can be used for expression of full length hTRT in *E. coli*. The expressed product will contain 29 amino acids encoded by the vector pBBS235, followed by 18 amino acids encoded by the 5'UTR of hTRT, followed by the full length hTRT protein.

In a further embodiment of the invention, *in vitro* mutagenesis of pGRN125 was done to convert the hTRT initiating ATG codon into a Kozak consensus and create EcoRI and BglII restriction digest sites to facilitate cloning into expression vectors. The oligonucleotide

5'-

TGCGCACGTGGGAAGCCCTGGCagatctgAattCcaCcATGCCGCGCGCTCCCCGC
TG-3' (altered nucleotides in lower case) was used in the mutagenesis procedure.

The resulting expression vector was designated pGRN127.

In another embodiment of the invention, the second Asp of the TRT "DD motif" was converted to an alanine to create a non-functional telomerase

enzyme, thus creating a mutant TRT protein for use as a dominant/negative mutant. The hTRT coding sequence was mutagenized *in vitro* using the oligonucleotide 5'-CGGGACGGGCTGCTCCTGCGTTTGGTGGAcGcgTTCTTGTTGGTGACACCTCACCTCACC-3' to convert the asparagine codon for residue 869 (Asp869) to an alanine (Ala) codon. This also created an MluI restriction enzyme site. The resulting expression plasmid was designated pGRN130, which also contains the Kozak consensus sequence as described for pGRN127.

The invention also provides a vector designed to express an antisense sequence fragment of hTRT. The pGRN126 plasmid was cut to completion with MscI and SmaI restriction enzymes and religated to delete over 95% of the hTRT ORF. One SmaI-MscI fragment was re-inserted during the process to recreate CAT activity. This unpurified plasmid was then redigested with SalI and EcoRI and the fragment containing the initiating codon of the hTRT ORF was inserted into the SalI-EcoRI sites of pBBS212 to make an antisense expression plasmid expressing the antisense sequence spanning the 5'UTR and 73 bases pair residues of the hTRT ORF (in mammalian cells). This plasmid was designated pGRN135.

Expression of hTRT Telomerase in Yeast

The present invention also provides hTRT-expressing yeast expression vectors to produce large quantities of full-length, biologically active hTRT.

Pichia pastoris Expression Vector pPICZ B and Full Length hTRT

To produce large quantities of full-length, biologically active hTRT, the *Pichia pastoris* expression vector pPICZ B (Invitrogen, San Diego, CA) was selected. The hTRT-coding sequence insert was derived from nucleotides 659 to 4801 of the hTRT insert in plasmid pGRN121. This nucleotide sequence includes the full-length sequence encoding hTRT. This expression vector is designed for inducible expression in *P. pastoris* of high levels of full-length, unmodified hTRT protein. Expression is driven by a yeast promoter, but the expressed sequence utilizes the hTRT initiation and termination codons. No exogenous codons were introduced by the cloning. The resulting pPICZ B/hTRT vector was used to transform the yeast.

Pichia pastoris Expression Vector hTRT-His6/pPICZ B

A second *Pichia pastoris* expression vector of the invention derived from pPICZ B, also contains the full-length sequence encoding hTRT derived from nucleotides 659 to 4801 of the hTRT insert in the plasmid pGRN121. This hTRT-His6/pPICZ B expression vector encodes full length hTRT protein fused at its C-terminus to the Myc epitope and His6 reporter tag sequences. The hTRT stop codon has been removed and replaced by vector sequences encoding the Myc epitope and the His6 reporter tag as well as a stop codon. This vector is designed to direct high-level inducible expression in yeast of the following fusion protein, which consists of hTRT sequence (underlined), vector sequences in brackets ([L] and [NSAVD]) the Myc epitope (double underlined), and the His6 tag (italicized):

MPRAPRCRAVRSLLRSHYREVLPLATFVRRLGPOGWRLVORGDP
AAFRALVAOCLVCVPWDARPPPAAPSFROVSKELVARVLQRLCERGAKNVL
15 AFGFALLDGARGGPPEAFTTSVRSYLPNTVTDALRGSGAWGLLLRRVGDDVLV
HLLARCALFVLVAPSCAYQVCGPPLYQLGAATQARPPPHASGPRRRLGCERAW
NHSVREAGVPLGLPAPGARRRGGSASRSLPLPKRPRRGAAPEPERTPVGQGSW
AHPGRTRGPSDRGFCVVSPARPAEEATSLEGALSGTRHSHPSVGROHHAGPPST
SRPPRPWDTPCPPVYAETKHFLYSSGDKEQLRPSFLLSSLRPSLTGARRLVETIFL
20 GSRPWMPGTPRRLPRLPQRYWQMRPLFLELLGNHAQCPYGVLLKTHCPLRAA
VTPAAGVCAREKPQGSVAAPEEEDTDPRLVQLLROHSSPWQVYGFVRACLR
LVPPGLWGSRHNERFLRNTKKFISLGKHAKLSLOELTWKMSVRDCAWLRRSP
GVGCVPAAEHRLREEILAKFLHWLMSVYVVELLSFFYVTETTFOKNRLFFYRK
SVWSKLOSIGIROHLKRVOLRELSEAEVROHREARPALLTSRLRFIPKPDGLRPI
25 VNMDYVVGARTFRREKRAERLTSRVKALFSVLNYERARRPGLLGASVLGLDDI
HRAWRTFVLRVRAODPPPELYFVKVDVTGAYDTIPODRLTEVIASIIKPONTYC
VRRYAVVQKAAHGHVRKAFKSHVSTLTDLOPYMROFVAHLOETSPLRDAVVI
EOSSSLNEASSGLFDVFLRFMCHHAVRIRGKSYVOCOGIPOGSILSTLLCSLCYG
DMENKLFAGIRRDGLLLRLVDDELLVTPHLTHAKTFLRTLVRGVPEYGCVVNL
30 RKTVVNFPVEDEALGGTAFVQMPAHGLFPWCGLLLDTRTLEVOSDYSSYARTS
IRASLTFNRGFKAGRNMRRLFGVLRLLKCHSLFLDLQVNSLOTVCTNIYKILLQ

AYRFHACVLQLPFHQQVWKNPTFFLRVISDTASLCYSILKAKNAGMSLGAKGA
AGPLPSEAVQWLCHOAFLKLTRHRVTYVPLLGSLRTAQTQLSRKLPGTTLTAL
EAAANPALPSDFKTILD[L]EOKLISEEDL[NSAVD]HHHHHH

5 Expression of hTERT in Insect Cells

The present invention also provides hTERT telomerase-expressing insect cell expression vectors that produce large quantities of full-length, biologically active hTERT.

10 Baculovirus Expression Vector pVL1393 and Full Length hTERT

The telomerase coding sequence of interest was cloned into the baculovirus expression vector pVL1393 (Invitrogen, San Diego, CA). This construct was subsequently cotransfected into *Spodoptera frugiperda* (sf-9) cells with linearized DNA from *Autographa californica* nuclear polyhedrosis virus (Baculogold-AcMNPV).

15 The recombinant baculoviruses obtained were subsequently plaque purified and expanded following standard protocols.

This expression vector provides for expression in insect cells of high levels of full-length hTERT protein. Expression is driven by a baculoviral polyhedrin gene promoter. No exogenous codons were introduced by the cloning.

20

Baculovirus Expression Vector pBlueBacHis2 B and Full Length hTERT

To produce large quantities of full-length, biologically active hTERT, the baculovirus expression vector pBlueBacHis2 B (Invitrogen, San Diego, CA) was selected as a source of control elements. The hTERT-coding insert consisted of
 25 nucleotides 707 to 4776 of the hTERT insert in plasmid pGRN121.

A full length hTERT with a His6 and Anti-Xpress tags (Invitrogen) was also constructed. This vector also contains an insert consisting of nucleotides 707 to 4776 of the hTERT insert from the plasmid pGRN121. The vector directs expression in insect cells of high levels of full length hTERT protein fused to a cleavable 6-histidine and

30 Anti-Xpress tags, and the amino acid sequence of the fusion protein is shown below;

(-*-*) denotes enterokinase cleavage site:

MPRGSHHHHHHGMASMTGGQQMGRDLYDDDDL-*DPSSRSAAGTME
 FAAASTQRCVLLRTWEALAPATPAMPRAAPRCRAVRSLLRSHYREVLPLATFVR
 RLGPQGWRLVQRGDPAAFRALVAQCLVCVPWDARPPPAAPSFRQVSCLKELV
 ARVLQRLCERGAKNVLAFGFALLDGARGGPPEAFTTSVRSYLPNTVTDALRGS
 5 GAWGLLLRRVGDVVLVHLLARCALFVLVAPSCAYQVCGPPLYQLGAATQARP
 PPHASGPRRLGCERAWNHSVREAGVPLGLPAPGARRRGGSASRSLPLPKRPRR
 GAAPEPERTPVGQGSWAHPGRTRGPSDRGFCVVSPARPAEEATSLEGALSGTR
 HSHPSVGRQHHAGPPSTSRPPRPWDTPCPPVYAETKHFLYSSGDKEQLRPSFLLS
 SLRPSLTGARRLVETIFLGSRPWMPGTPRRLPRLPQRYWQMRPLFLELLGNHAQ
 10 CPYGVLKTHCPLRAAVTPAAGVCAREKPQGSVAAPEEEDTDPRRLVQLLROH
 SSPWQVYGFVRACLRLVPPGLWGSRHNERFLRNTKKFISLGKHAKLSLQELT
 WKMSVRDCAWLRRSPGVGCVPAAEHRLREEILAKFLHWLMSVYVVELLSFF
 YVTETTFQKNRLFFYRKS VWSKLQSIGIRQHLKRVQLRELSEAEVRQHREARPA
 LLTSRLRFIPKPDGLRPIVNMDYVVGARTFRREKRAERLTSRVKALFSVLNYER
 15 ARRPGLLGASVLGLDDIHRWRTFVLRVRAQDPPPELYFVKVDVTGAYDTIPQ
 DRLTEVIASIIKPQNTYCVRRYAVVQKAAHGHVRKAFKSHVSTLTDLPYMRQ
 FVAHLQETSPLRDVAVVIEQSSSLNEASSGLFDVFLRFMCHHAVRIRGKSYVQCQ
 GIPQGSILSTLLCSLCYGD MENKLFAGIRRDGLLLRLVDDFLLVTPHLTHAKTFL
 RTLVRGVPEYGCVVNLKRTVNFVPEDEALGGTAFVQMPAHGLFPWCGLLLD
 20 TRTLEVQSDYSSYARTSIRASLTFNRGFKAGRNMRRKLFGVLRLLKCHSLFLDLQ
 VNSLQTVCTNIYKILLQAYRFHACVLQLPFHQQVWKNPTFFLRVISDTASLCY
 SILKAKNAGMSLGAKGAAGPLPSEAVQWLCHQAFLKLTRHRVTYVPLLGLSLR
 TAQTQLSRKLP GTTLTALEAAANPALPSDFKTILD

25 **Baculovirus Expression Vector *pBlueBac4.5* and Full Length hTRT Protein**

To produce large quantities of full-length, biologically active hTRT, a second baculovirus expression vector, *pBlueBac4.5* (Invitrogen, San Diego, CA) was constructed. The hTRT-coding insert also consisted of nucleotides 707 to 4776 of the hTRT from the plasmid *pGRN121*.

30

Baculovirus Expression Vector *pMelBacB* and Full Length hTRT Protein

To produce large quantities of full-length, biologically active hTRT, a third baculovirus expression vector, *pMelBacB* (Invitrogen, San Diego, CA) was constructed. The hTRT-coding insert also consists of nucleotides 707 to 4776 of the hTRT insert from the plasmid *pGRN121*.

35

pMelBacB directs expression of full length hTRT in insect cells to the extracellular medium through the secretory pathway using the melittin signal sequence. High levels of full length hTRT are thus secreted. The melittin signal sequence is cleaved upon excretion, but is part of the protein pool that remains intracellularly. For

that reason, it is indicated in parentheses in the following sequence. The sequence of the fusion protein encoded by the vector is shown below:

(MKFLVNVALVFMVVYISYIYA)-*-DPSSRSAAGTMEFAAASTQRCVLLR
TWEALAPATPAMPRAAPRCRAVRSLLRSHYREVLPLATFVRRLGPQGWRLVQRG
5 DPAAFRALVAQCLVCVPWDARPPPAAPSFRQVSCLELVARVLQRLCERGAKN
VLAFGFALLDGARGGPPEAFTTSVRSYLPNTVTDALRGSGAWGLLLRRVGDDV
LVHLLARCALFVLVAPSCAYQVCGPPLYQLGAATQARPPPHASGPRRLGCER
AWNHSVREAGVPLGLPAPGARRRGGSASRSLPLPKRPRRGAAPEPERTPVGQG
SWAHPGRTRGPSDRGFCVVSPARPAEEATSLEGALSGTRHSHPSVGRQHHAGP
10 PSTSRPPRPWDTPCPPVYAETKHFLYSSGDKEQLRPSFLLSSLRPSLTGARRLVE
TIFLGSRPWMPGTPRRLPRLPQRYWQMRPLFLELLGNHAQCPYGVLLKTHCPL
RAAVTPAAGVCAREKPQGSVAAPEEEDTDPRRLVQLLRQHSSPWQVYGFVRA
CLRRLVPPGLWGSRHNERFLRNTKKFISLGKHAKLSLQELTWKMSVRDCAWL
RRSPGVGCVPAAEHLREEILAKFLHWLMSVYVVELLRSSFFYVTETTFQKNRLF
15 FYRKS VWSKLQSIGIRQHLKRVQLRELSEAEVRQHREARPALLTSRLRFIPKPDG
LRPIVNMDYVVGARTFRREKRAERLTSRVKALFSVLNYERARRPGLLGASVLG
LDDIHRWRTFVLRVRAQDPPPELYFVKVDVTGAYDTIPQDRLTEVIASIIKPQN
TYCVRRYAVVQKAAHGHVRKAFKSHVSTLTDLQPYMRQFVAHLQETSPLRDA
VVIEQSSSLNEASSGLFDVFLRFMCHHAVRIRGKSYVQCQGIQGSILSTLLCSLC
20 YGDMENKLFAGIRRDGLLLRLVDDFLLVTPHLTHAKTFLRTLVRGVPEYGCVV
NLRKTVVNFPVEDEALGGTAFVQMPAHGLFPWCGLLLDTRTLEVQSDYSSYAR
TSIRASLTFNRGFKAGRNMRRKLFGLRLKCHSLFLDLQVNSLQTVCTNIYKILL
LQAYRFHACVLQLPFHQVWKNPTFFLRVISDTASLCYSILKAKNAGMSLGAK
GAAGPLPSEAVQWLCHQAFLKLTRHRVTYVPLLGLSLRTAQTQLSRKLPGTTL
25 TALEAAANPALPSDFKTILD

Expression of hTERT in Mammalian Cells

The present invention also provides vectors to produce hTERT in large quantities as full-length, biologically active protein in a variety of mammalian cell
30 lines, which is useful in many embodiments of the invention, as discussed above.

MPSV-hTERT Expression Plasmids

The invention also provides for an expression system for use in mammalian cells that gives the highest possible expression of recombinant protein,
35 such as telomerase, without actually modifying the coding sequence (e.g. optimizing codon usage). In one embodiment, the invention provides MPSV mammalian expression plasmids (from plasmid pBBS212, described as pMPSV-TM in Lin J-H (1994) *Gene* 47:287-292) capable of expressing the TRTs of the invention. The MPSV-plasmids can be expressed either as stable or transient clones.

In this expression system, while the hTRT coding sequence itself is unchanged, exogenous transcriptional control elements are incorporated into the vector. The myeloproliferative sarcoma virus (MPSV) LTR (MPSV-LTR) promoter, enhanced by the cytomegalovirus (CMV) enhancer, is incorporated for transcriptional initiation. This promoter consistently shows higher expression levels in cell lines (see Lin J-H (1994) *supra*). A Kozak consensus sequence can be incorporated for translation initiation (see Kozak (1996) *Mamm. Genome* 7:563-574). All extraneous 5' and 3' untranslated hTRT sequences can be removed to insure that these sequences do not interfere with expression, as discussed above. The MPSV plasmid containing the complete hTRT coding sequence, with all extraneous sequences included, is designated pGRN133. A control, hTRT "antisense" plasmid was also constructed. This vector is identical to pGRN133 except that the TRT insert is the antisense sequence of hTRT (the antisense, which control can be used as a vector is designated pGRN134). The MPSV plasmid containing the complete hTRT coding sequence with all other extraneous sequences removed and containing the Kozak consensus sequence is designated pGRN145.

Two selection markers, PAC (Puromycin-N-acetyl-transferase = Puromycin resistance) and HygB (Hygromycin B = Hygromycin resistance) are present for selection of the plasmids after transfection (see discussion referring to selectable markers, above). Double selection using markers on both sides of the vector polylinker should increase the stability of the hTRT coding sequence. A DHFR (dihydrofolate reductase) encoding sequence is included to allow amplification of the expression cassette after stable clones are made. Other means of gene amplification can also be used to increase recombinant protein yields.

The invention also provides for MPSV mammalian expression plasmids containing hTRT fusion proteins. In one embodiment, the hTRT sequence, while retaining its 5' untranslated region, is linked to an epitope flag, such as the IBI FLAG (International Biotechnologies Inc. (IBI), Kodak, New Haven, CT) and inserted into the MPSV expression plasmid (designated pGRN147). This particular construct contains a Kozak translation initiation site. The expressed fusion protein can be purified using the M-1 anti-FLAG octapeptide monoclonal antibody (IBI, Kodak,

supra).

In another embodiment, hTRT is site-specifically altered. One amino acid residue codon is mutagenized, changing the aspartic acid at position 869 to an alanine. This Asp869->Ala hTRT mutant, retaining its 5' untranslated region and incorporating a Kozak sequence, was inserted into an MPSV expression plasmid, and designated pGRN146. The Asp869->Ala hTRT mutant was further engineered to contain the FLAG sequence, as described above, and the insert cloned into an MPSV expression plasmid. This expression plasmid is designated pGRN154. Specifically, for pGRN154, an Eam1105I restriction digest fragment from pGRN146 containing the Kozak sequence-containing "front end" (5' segment) of hTRT is cloned into the Eam1105I sites of pGRN147 (see above) to make an MPSV expression plasmid capable of expressing hTRT with a Kozak sequence, the above-described D869->A mutation, and the IBI flag.

Another embodiment of the invention is an expression plasmid derived from pGRN146. The mammalian expression plasmid, designated pGRN152, was generated by excising the EcoRI fragment from plasmid pGRN146 (containing the hTRT ORF) and cloned into the EcoRI site of pBBS212 to remove the 5'UTR of hTRT. The hTRT is oriented so that its expression is controlled by the MPSV promoter. This makes a mammalian expression plasmid that expresses hTRT with a Kozak consensus sequence and the D869->A mutation, and uses the MPSV promoter.

The invention provides for a mammalian expression vector in which hTRT is oriented so that the hTRT coding sequence is driven by the MPSV promoter. For example, an EcoRI restriction digest fragment from pGRN137 containing the hTRT open reading frame (ORF) was cloned into the EcoRI site of pBBS212 (see below), thus removing the 5' untranslated region (5'-UTR) of hTRT. pGRN137 was constructed by excising a SalI-Sse8387I fragment from pGRN130, described below, containing the Kozak mutation of hTRT into the Sal I-SSE 8387I sites of pGRN136, making a mammalian expression plasmid expressing hTRT containing a Kozak consensus sequence off the MPSV promoter. Plasmid pGRN136 was constructed by excising a HindIII SalI fragment from pGRN126 containing the

hTERT ORF and cloning it into the HindIII SalI sites of plasmid, pBBS242, making a mammalian expression plasmid expressing hTERT off the MPSV promoter). This makes a mammalian expression plasmid, designated pGRN145, that expresses hTERT with a Kozak consensus sequence using the MPSV promoter. See also the pGRN152
5 MPSV promoter-driven mammalian expression vector described below.

hTERT Expressed in 293 Cells using Episomal Vector pEBVHis

An episomal vector, pEBVHis (Invitrogen, San Diego, CA) was engineered to express an hTERT fusion protein comprising hTERT fused to an
10 N-terminal extension epitope tag, the Xpress epitope (Invitrogen, San Diego, CA) (designated pGRN122). The NotI hTERT fragment from pGRN121 containing the hTERT ORF was cloned into the NotI site of pEBVHisA so that the hTERT ORF is in the same orientation as the vector's Rous Sarcoma Virus (RSV) promoter. In this orientation the His6 flag was relatively closer to the N-terminus of hTERT.

15 A vector was also constructed containing as an insert the antisense sequence of hTERT and the epitope tag (the plasmid designated pGRN123, which can be used as a control). The vector was transfected into 293 cells and translated hTERT identified and isolated using an antibody specific for the Xpress epitope. pEBVHis is a hygromycin resistant EBV episomal vector that expresses the protein of interest
20 fused to a N-terminal peptide. Cells carrying the vector are selected and expanded, then nuclear and cytoplasmic extracts prepared. These and control extracts are immunoprecipitated with anti-Xpress antibody, and the immunoprecipitated beads are tested for telomerase activity by conventional assay.

25 Expression Recombinant hTERT in Mortal, Normal Diploid Human Cells

In one embodiment of the invention, recombinant hTERT and necessary telomerase enzyme complex components can be expressed in normal, diploid mortal cells to increase their proliferative capacity or to immortalize them, or to facilitate immortalizing them. This allows one to obtain diploid immortal cells with an
30 otherwise normal phenotype and karyotype. As discussed above, this use of telomerase has enormous commercial utility.

Sense hTERT (Figure 16) and antisense hTERT) were cloned into a CMV vector. These vectors were purified and transiently transfected into two normal, mortal, diploid human cell clones. The human clones were young passage diploid human BJ and IMR90 cell strains.

5 Analysis of telomerase activity using a TRAP assay (utilizing the TRAPeze™ Kit (Oncor, Inc., Gaithersburg, MD) showed that transfection of sense hTERT - but not antisense hTERT - generated telomerase activity in both the BJ and IMR90 cell strains.

10 **Expression of Recombinant hTERT in Immortalized IMR90 Human Cells**

 Using the same hTERT sense construct cloned into CMV vectors used in the above described diploid human BJ and IMR90 cell strains studies, immortalized SW13 ALT pathway cell line (an IMR90 cell immortalized with SV40 antigen) was transiently transfected. A TRAP assay (TRAPeze, Oncor, Inc, Gaithersburg, MD) demonstrated that telomerase activity was generated in the sense construct transfected cells.

Vectors for Regulated Expression of hTERT in Mammalian Cells: Inducible and Repressible Expression of hTERT

20 The invention provides vectors that can be manipulated to induce or repress the expression of the TRTs of the invention, such as hTERT. For example, the hTERT coding sequence can be cloned into the Ecdysone-Inducible Expression System from Invitrogen (San Diego, CA) and the Tet-On and Tet-off tetracycline regulated systems from Clontech Laboratories, Inc. (Palo Alto, CA). Such inducible
25 expression systems are provided for use in the methods of the invention where it is important to control the level or rate of transcription of transfected TRT. For example, the invention provides for cell lines immortalized through the expression of hTERT; such cells can be rendered "mortal" by inhibition of hTERT expression by the vector through transcriptional controls, such as those provided by the Tet-Off system.
30 The invention also provides for methods of expressing TRT only transiently to avoid the constitutive expression of hTERT, which may lead to unwanted "immortalization"

of the transfected cells, as discussed above.

The Ecdysone-Inducible Mammalian Expression System is designed to allow regulated expression of the gene of interest in mammalian cells. The system is distinguished by its tightly regulated mechanism that allows almost no detectable basal expression and greater than 200-fold inducibility in mammalian cells. The expression system is based on the heterodimeric ecdysone receptor of *Drosophila*. The Ecdysone-Inducible Expression System uses a steroid hormone ecdysone analog, muristerone A, to activate expression of hTRT via a heterodimeric nuclear receptor. Expression levels have been reported to exceed 200-fold over basal levels with no effect on mammalian cell physiology "Ecdysone-Inducible Gene Expression in Mammalian Cells and Transgenic Mice" (1996) *Proc. Natl. Acad. Sci. USA* 93, 3346-3351). Once the receptor binds ecdysone or muristerone, an analog of ecdysone, the receptor activates an ecdysone-responsive promoter to give controlled expression of the gene of interest. In the Ecdysone-Inducible Mammalian Expression System, both monomers of the heterodimeric receptor are constitutively expressed from the same vector, pVgRXR. The ecdysone-responsive promoter, which ultimately drives expression of the gene of interest, is located on a second vector, pIND, which drives the transcription of the gene of interest.

The hTRT coding sequence is cloned in the pIND vector (Clontech Laboratories, Inc, Palo Alto, CA), which contains 5 modified ecdysone response elements (E/GREs) upstream of a minimal heat shock promoter and the multiple cloning site. The construct is then transfected in cell lines which have been pre-engineered to stably express the ecdysone receptor. After transfection, cells are treated with muristerone A to induce intracellular expression from pIND.

The Tet-on and Tet-off expression systems (Clontech, Palo Alto, CA) give access to the regulated, high-level gene expression systems described by Gossen (1992) "Tight control of gene expression in mammalian cells by tetracycline responsive promoters" *Proc. Natl. Acad. Sci. USA* 89:5547-5551, for the Tet-Off transcription repression system; and Gossen (1995) "Transcriptional activation by tetracycline in mammalian cells" *Science* 268:1766-1769, for the Tet-On inducible transcriptional system. In "Tet-Off" transformed cell lines, gene expression is turned on when

tetracycline (Tc) or doxycycline ("Dox;" a Tc derivative) is removed from the culture medium. In contrast, expression is turned on in Tet-On cell lines by the addition of Tc or Dox to the medium. Both systems permit expression of cloned genes to be regulated closely in response to varying concentrations of Tc or Dox.

5 This system uses the "pTRE" as a response plasmid that can be used to express a gene of interest. Plasmid pTRE contains a multiple cloning site (MCS) immediately downstream of the Tet-responsive PhCMV*-1 promoter. Genes or cDNAs of interest inserted into one of the sites in the MCS will be responsive to the tTA and rtTA regulatory proteins in the Tet-Off and Tet-On systems, respectively. PhCMV*-1
10 contains the Tet-responsive element (TRE), which consists of seven copies of the 42-bp tet operator sequence (tetO). The TRE element is just upstream of the minimal CMV promoter (PminCMV), which lacks the enhancer that is part of the complete CMV promoter in the pTet plasmids. Consequently, PhCMV*-1 is silent in the absence of binding of regulatory proteins to the tetO sequences. The cloned insert must have an
15 initiation codon. In some cases, addition of a Kozak consensus ribosome binding site may improve expression levels; however, many cDNAs have been efficiently expressed in Tet systems without the addition of a Kozak sequence. pTRE-Gene X plasmids are cotransfected with pTK-Hyg to permit selection of stable transfectants.

 Setting up a Tet-Off or Tet-On expression system generally requires two
20 consecutive stable transfections to create a "double-stable" cell line that contains integrated copies of genes encoding the appropriate regulatory protein and TRT under the control of a TRE. In the first transfection, the appropriate regulatory protein is introduced into the cell line of choice by transfection of a "regulator plasmid" such as pTet-Off or pTet-On vector, which expresses the appropriate regulatory proteins. The
25 hTRT cloned in the pTRE "response plasmid" is then introduced in the second transfection to create the double-stable Tet-Off or Tet-On cell line. Both systems give very tight on/off control of gene expression, regulated dose-dependent induction, and high absolute levels of gene expression.

30 **Expression Recombinant hTRT With DHFR and Adenovirus Sequences**

 The pGRN155 plasmid construct was designed for transient expression

of hTRT cDNA in mammalian cells. A Kozak consensus is inserted at the 5' end of the hTRT sequence. The hTRT insert contains no 3' or 5' UTR. The hTRT cDNA is inserted into the EcoRI site of p91023(B) (Wong (1985) *Science* 228:810-815). The hTRT insert is in the same orientation as the DHFR ORF. This makes the expression vector particularly useful for transient expression.

Plasmid pGRN155 contains the SV40 origin and enhancer just upstream of an adenovirus promoter, a tetracycline resistance gene, an *E. coli* origin and an adenovirus VAI and VAI gene region. This expression cassette contains, in the following order: the adenovirus major late promoter; the adenovirus tripartite leader; a hybrid intron consisting of a 5' splice site from the first exon of the tripartite leader and a 3' splice site from the mouse immunoglobulin gene; the hTRT cDNA; the mouse DHFR coding sequence; and, the SV40 polyadenylation signal.

The adenovirus tripartite leader and the VA RNAs have been reported to increase the efficiency with which polycistronic mRNAs are translated. DHFR sequences have been reported to enhance the stability of hybrid mRNA. DHFR sequences also can provide a marker for selection and amplification of vector sequences. See Logan (1984) *Proc. Natl. Acad. Sci. USA* 81:3655; Kaufman (1985) *Proc. Natl. Acad. Sci. USA* 82: 689 ; and Kaufman (1988) *Focus* (Life Technologies, Inc.), Vol.10, no. 3).

Other expression plasmids are described for illustrative purposes.

pGRN121

The EcoRI fragment from lambda clone 25-1.1.6 containing the entire cDNA encoding hTRT protein was inserted into the EcoRI site of pBluescriptIISK+ such that the 5' end of the cDNA is near the T7 promoter in the vector. The selectable marker that is used with this vector is ampicillin.

pGRN122

The NotI fragment from pGRN121 containing the hTRT ORF was inserted into the NotI site of pEBVHisA so that the coding sequence is operably linked to the RSV promoter. This plasmid expresses a fusion protein composed of a His6 flag fused to the

N-terminal of the hTRT protein. The selectable marker that is used with this vector is ampicillin or hygromycin.

pGRN123

- 5 The NotI fragment from pGRN121 containing the hTRT ORF was inserted into the NotI site of pEBVHisA so that the coding sequence is in the opposite orientation as the RSV promoter, thus expressing *antisense* hTRT.

pGRN124

- 10 Plasmid pGRN121 was deleted of all ApaI sites followed by deletion of the MscI-HincII fragment containing the 3'UTR. The Nco-XbaI fragment containing the stop codon of the hTRT coding sequence was then inserted into the Nco-XbaI sites of pGRN121 to make a plasmid equivalent to pGRN121 except lacking the 3'UTR, which may be preferred for increased expression levels in some cells.

15

pGRN125

- The NotI fragment from pGRN124 containing the hTRT coding sequence was inserted into the NotI site of pBBS235 so that the open reading frame is in the opposite orientation of the Lac promoter. The selectable marker that is used with this vector is chloramphenicol.
- 20

pGRN126

- The NotI fragment from pGRN124 containing the hTRT coding sequence was inserted into the NotI site of pBBS235 so that the hTRT coding sequence inserted is in the same orientation as the Lac promoter.
- 25

pGRN127

- The oligonucleotide 5'-TGCGCACGTGGGAAGCCCTGGCagatctgAattCCaCcATGC CGCGCGCTCCCCGCTG-3' was used in *in vitro* mutagenesis of pGRN125 to convert the initiating ATG codon of the hTRT coding sequence into a Kozak consensus sequence and create EcoRI and BglII sites for cloning. Also, oligonucleotide COD2866
- 30

was used to convert AmpS to AmpR (ampicillin resistant) and oligonucleotide COD1941 was used to convert CatR (chloramphenicol resistant) to CatS (chloramphenicol sensitive).

5 pGRN128

The oligonucleotide 5'-TGCGCACGTGGGAAGCCCTGGCagatctgAattCCaCcATG CCGCGCGCTCCCCGCTG-3' is used in *in vitro* mutagenesis to convert the initiating ATG codon of hTRT into a Kozak consensus and create EcoRI and BglII sites for cloning. Also, oligo 5'-

- 10 CTGCCCTCAGACTTCAAGACCATCCTGGACTACAA
GGACGACGATGACAAATGAATTCAGATCTGCGGCCGCCACCGCGGTGGAG
CTCCAGC-3' is used to insert the IBI Flag (International Biotechnologies Inc. (IBI), Kodak, New Haven, CT) at the C-terminus and create EcoRI and BglII sites for cloning. Also, COD2866 is used to convert AmpS to AmpR and COD1941 is
15 used to convert CatR to CatS.

pGRN129

The oligonucleotide 5'-

- CGGGACGGGCTGCTCCTGCGTTTGGTGGAcGcgTTCTTG
20 TTGGTGACACCTCACCTCACC-3' was used by *in vitro* mutagenesis to convert Asp869 to an Ala codon (i.e. the second Asp of the DD motif was converted to an Alanine to create a dominant/negative hTRT mutant). This also created a MluI site. Also, oligonucleotide 5'-
CTGCCCTCAGACTTCAAGACCATCCTGGACTACAAGG
25 ACGACGATGACAAATGAATTCAGATCTGCGGCCGCCACCGCGGTGGAGCT
CCAGC-3') was used to insert the IBI Flag at the C-terminus and create EcoRI and BglII sites for cloning. Also, COD2866 was used to convert AmpS to AmpR and COD1941 was used to convert CatR to CatS.

30 pGRN130

The oligonucleotide 5'-CGGGACGGGCTGCTCCTGCGTTTGGTGGAcGcgTTCTT

GTTGGTGACACCTCACCTCACC-3' was used in *in vitro* mutagenesis to convert the Asp869 codon into an Ala codon (i.e. the second Asp of the DD motif was converted to an Alanine to make a dominant/negative variant protein). This also created an MluI site.

- Also, the oligonucleotide 5'-TGCGCACGTGGGAAGCCCTGGCagatctgAatt
5 CcCaATGCCGCGCGCTCCCCGCTG-3' was used in *in vitro* mutagenesis to convert the initiating ATG codon of the hTRT coding sequence into a Kozak consensus sequence and create EcoRI and BglII sites for cloning. Also, COD2866 was used to convert AmpS to AmpR and COD1941 was used to convert CatR.

10 **pGRN131 (not of the invention)**

The EcoRI fragment from pGRN128 containing the hTRT ORF with Kozak sequence and IBI Flag mutations is inserted into the EcoRI site of pBBS212 so that the hTRT ORF is expressed off the MPSV promoter. Plasmid pBBS212 contains a MPSV promoter, the CMV enhancer, and the SV40 polyadenylation site.

15

pGRN132(not of the invention)

The EcoRI fragment from pGRN128 containing the hTRT ORF with Kozak sequence and IBI Flag mutations is inserted into the EcoRI site of pBBS212 so that the antisense of the hTRT ORF is expressed off the MPSV promoter.

20

pGRN133

The EcoRI fragment from pGRN121 containing the hTRT coding sequence was inserted into the EcoRI site of pBBS212 so that the hTRT protein is expressed under the control of the MPSV promoter.

25

pGRN134

The EcoRI fragment from pGRN121 containing the hTRT coding sequence was inserted into the EcoRI site of pBBS212 so that the *antisense* of the hTRT coding sequence is expressed under the control of the MPSV promoter. The selectable markers used with

- 30 this vector are Chlor/HygB/PAC.

pGRN135

Plasmid pGRN126 was digested to completion with MscI and SmaI and religated to delete over 95% of the hTRT coding sequence inserted. One SmaI-MscI fragment was re-inserted during the process to recreate the Cat activity for selection. This unpurified
5 plasmid was then redigested with Sall and EcoRI and the fragment containing the initiating codon of the hTRT coding sequence was inserted into the Sall-EcoRI sites of pBBS212. This makes an *antisense* expression plasmid expressing the antisense of the 5'UTR and 73 bases of the coding sequence. The selectable markers used with this vector are Chlor/HygB/PAC.

10

pGRN136

The HindIII-Sall fragment from pGRN126 containing the hTRT coding sequence was inserted into the HindIII-Sall sites of pBBS242.

pGRN137

The Sall-Sse8387I fragment from pGRN130 containing the Kozak sequence was inserted into the Sall-Sse8387I sites of pGRN136.

pGRN138

20 The EcoRI fragment from pGRN124 containing hTRT minus the 3'UTR was inserted into the EcoRI site of pEGFP-C2 such that the orientation of the hTRT is the same as the EGFP domain.

pGRN139

25 The oligonucleotide 5'- CTGCCCTCAGACTTCAAGACCATCCTGGACTACAAGG ACGACGATGACAAATGAATTCAGATCTGCGGCCGCCACCGCGGTGGAGCTC CAGC-3 was used in *in vitro* mutagenesis to insert the IBI Flag at the C-terminus of hTRT in pGRN125 and create EcoRI and BglII sites for cloning. Also, COD2866 was used to convert AmpS to AmpR) and COD1941 was used to convert CatR to CatS.

30

pGRN140

The NcoI fragment containing the upstream sequences of genomic hTRT and the first intron of hTRT from lambdaG55 was inserted into the NcoI site of pBBS167. The fragment is oriented so that hTRT is in the same direction as the Lac promoter.

5 **pGRN141**

The NcoI fragment containing the upstream sequences of genomic hTRT and the first intron of hTRT from lambdaG55 was inserted into the NcoI site of pBBS167. The fragment is oriented so that hTRT is in the opposite direction as the Lac promoter.

10 **pGRN142**

The NotI fragment from lambdaGphi5 containing the complete ~15 kbp genomic insert including the hTRT gene promoter region was inserted in the NotI site of plasmid pBBS185. The fragment is oriented so that the hTRT ORF is in the opposite orientation as the Lac promoter.

15

pGRN143

The NotI fragment from lambdaGphi5 containing the complete ~15 kbp genomic insert including the hTRT gene promoter region was inserted in the NotI site of plasmid pBBS185. The fragment is oriented so that the hTRT ORF is in the same orientation as

20 the Lac promoter.

pGRN144

SAL1 deletion of pGRN140 to remove lambda sequences.

25 **pGRN145 (not of the invention)**

This vector was an intermediate vector for constructing a hTRT fusion protein expression vector. The EcoRI fragment from pGRN137 containing the hTRT coding sequence was inserted into the EcoRI site of pBBS212 to remove the portion of the sequence corresponding to the 5'UTR of hTRT mRNA. The hTRT coding sequence is
30 oriented so that it is expressed under the control of the MPSV promoter. The selectable markers used with this vector are Chlor/HygB/PAC.

pGRN146

This vector was constructed for the expression of hTRT sequences in mammalian cells. The Sse8387I-NotI fragment from pGRN130 containing the D869A mutation of hTRT
5 was inserted into the Sse8387I-NotI sites of pGRN137. The selectable markers used with this vector are Ampicillin/HygB/PAC.

pGRN147

The Sse8387I-NotI fragment from pGRN139 containing the IBI Flag was inserted into
10 the Sse8387I-NotI sites of pGRN137.

pGRN148

The BglII-Eco47III fragment from pGRN144 containing the promoter region of hTRT was inserted into the BglII-NruI sites of pSEAP2 to make an hTRT
15 promoter/reporter construct.

pGRN149

This vector was constructed for the expression and mutagenesis of TRT sequences in *E. coli*. The mutagenic oligo 5'-
20 cttcaagaccatcctggactttcgaaacgcggccgccaccgcggtggagctcc-3' was used to add a CSP45I site at the C-terminus of hTRT by *in vitro* mutagenesis of pGRN125. The "stop" codon of hTRT was deleted and replaced with a Csp45I site. The selectable marker that is used with this vector is ampicillin.

pGRN150

The BglII-FspI fragment from pGRN144 containing the promoter region of hTRT was inserted into the BglII-NruI sites of pSEAP2 to make an hTRT promoter/reporter construct.

pGRN151 (not of the invention)

This vector was constructed for the expression of hTRT sequences in mammalian cells.

The EcoRI fragment from pGRN147 containing the hTERT coding sequence was inserted into the EcoRI site of pBBS212 to remove the portion of the sequence corresponding to the 5'UTR of the hTERT mRNA. The hTERT coding sequence is oriented so that it is expressed under the control of the MPSV promoter. The selectable markers used with
5 this vector are Chlor/HygB/PAC.

pGRN152 (not of the invention)

The EcoRI fragment from pGRN146 containing the hTERT coding sequence was inserted into the EcoRI site of pBBS212 to remove the portion of the sequence corresponding to
10 the 5'UTR of the hTERT. The hTERT coding sequence is oriented so that it is expressed under the control of the MPSV promoter.

pGRN153 (not of the invention)

The StyI fragment from pGRN130 containing the D869-->A mutation of hTERT (hTERT
15 variant coding sequence) was inserted into the StyI sites of pGRN158 to make a plasmid containing the hTERT coding sequence with a Kozak consensus sequence at its 5'-end, an IBI FLAG sequence at its 3'-end (the C-terminus encoding region), and the D869-->A mutation.

20 pGRN154 (not of the invention)

The EcoRI fragment of pGRN153 containing the hTERT gene was inserted into the EcoRI site of plasmid pBS212 in an orientation such that the hTERT ORF is oriented in the same direction as the MPSV promoter. This makes an MPSV-directed expression plasmid that expresses the hTERT protein with a Kozak consensus sequence at its
25 amino-terminal end, an IBI FLAG at its carboxy-terminal end, and the D869-->A mutation

pGRN155 (not of the invention)

This vector was constructed for the expression of hTERT sequences in mammalian cells.
30 The insert included full length cDNA of hTERT minus 5' and 3' UTR, and Kozak sequences. The EcoRI fragment from pGRN145 containing the hTERT cDNA with the

Kozak consensus and no 3' or 5' UTR was inserted into the EcoRI site of p91023(B) such that the hTRT is in the same orientation as the DHFR ORF. This makes a transient expression vector for hTRT. The selectable marker used with this vector is tetracycline.

5

pGRN156 (not of the invention)

This vector was constructed for the expression of hTRT sequences in mammalian cells. The EcoRI fragment from pGRN146 containing the D869A mutation of the hTRT cDNA with the Kozak consensus and no 3' or 5' UTR was inserted into the EcoRI
10 site of p91023(B) such that the hTRT is in the same orientation as the DHFR ORF. This makes a transient expression vector for hTRT. The insert included full length cDNA of hTRT minus 5' and 3' UTR, D869A, and Kozak sequences. The selectable marker used with this vector is tetracycline.

15 **pGRN157 (not of the invention)**

This vector was constructed for the expression of hTRT sequences in mammalian cells. The EcoRI fragment from pGRN147 containing the hTRT cDNA with the IBI FLAG at the C-terminus; the Kozak consensus and no 3' or 5' UTR into the EcoRI site of p91023(B) such that the hTRT is in the same orientation as the DHFR ORF. This
20 makes a transient expression vector for hTRT. The insert included full length cDNA of hTRT minus 5' and 3' UTR, the FLAG sequence, and Kozak sequences. The selectable marker used with this vector is tetracycline.

pGRN158 (not of the invention)

25 This vector was constructed for the expression and mutagenesis of TRT sequences in *E. coli*. The EcoRI fragment from pGRN151 containing the hTRT ORF was inserted into the EcoRI site of pBBS183 so that the hTRT ORF is oriented in the opposite direction as the Lac promoter. The insert included full length cDNA of hTRT minus 5' and 3' UTR, FLAG sequence, and Kozak sequences. The hTRT coding sequence is
30 driven by a T7 promoter. The selectable marker used with this vector is ampicillin.

pGRN159

- This vector was constructed for the expression and mutagenesis of TRT sequences in *E. coli*. The Nhe-KpnI fragment from pGRN138 containing the EGFP to hTRT fusion was inserted into the XbaI-KpnI sites of pBluescriptIIKS+. This makes a T7
5 expression vector for the fusion protein (the coding sequence is driven by a T7 promoter). The insert included full length cDNA of hTRT minus the 3' UTR as a fusion protein with EGFP. The selectable marker used with this vector is ampicillin.

pGRN160 (not of the invention)

- 10 This vector was constructed for the expression of *antisense* hTR sequences in mammalian cells. The coding sequence is operably linked to an MPSV promoter. The XhoI-NsiI fragment from pGRN90 containing the full length hTR ORF was inserted into the SalI-Sse8387I sites of pBBS295. This makes a transient/stable vector expressing hTR antisense RNA. A GPT marker was incorporated into the vector.
15 The selectable markers used with this vector are Chlor/gpt/PAC.

pGRN161 (not of the invention)

- This vector was constructed for the expression of *sense* hTR sequences in mammalian cells. The XhoI-NsiI fragment from pGRN89 containing the full length hTR ORF
20 was inserted into the SalI-Sse8387I sites of pBBS295. This makes a transient/stable vector expressing hTR in the sense orientation. The coding sequence is driven by an MPSV promoter. A GPT marker was incorporated into the vector. The selectable markers used with this vector are Chlor/gpt/PAC.

pGRN162 (not of the invention)

- The XhoI-NsiI fragment from pGRN87 containing the full length hTR ORF was inserted into the SalI-Sse8387I sites of pBBS295. This makes a transient/stable vector expressing truncated hTR (from position +108 to +435) in the sense orientation.

pGRN163

- 30 This vector was constructed for the expression and mutagenesis of TRT sequences in *E.*

coli. The coding sequence is driven by a T7 promoter. Oligonucleotide RA45 (5'-GCCACCCCGCGCTGCCTCGAGCTCCCGCTGC-3') is used in *in vitro* mutagenesis to change the initiating met in hTRT to Leu and introduce an XhoI site in the next two codons after the Leu. Also COD 1941 was used to change CatR to CatS, and introduces a BSPH1 site, and COD 2866 was used to change AmpS to AmpR, introducing an FSP1 site. The selectable marker used with this vector is ampicillin.

pGRN164 (not of the invention)

10 This vector was constructed for the expression of hTR sequences in *E. coli*. Primers hTR+1 5'-GGGGAAGCTTTAATACGACTCACTATAGGGTTGCGGAGGGTGG GCCTG-3' and hTR+445 5'-CCCCGGATCCTGCGCATGTGTGAGCCGAGTCCT GGG-3' were used to amplify by PCR a fragment from pGRN33 containing the full length hTR with the T7 promoter on the 5' end (as in hTR+1). A BamHI-HindIII
15 digest of the PCR product was put into the BamHI-HindIII sites of pUC119. The coding sequence operably linked to a T7 promoter. The selectable marker used with this vector is ampicillin.

pGRN165 (not of the invention)

20 This vector was constructed for the expression and mutagenesis of hTRT sequences in *E. coli*. The coding sequence is operably linked to a T7 promoter. The EcoRI fragment from pGRN145 containing the hTRT ORF with a Kozak front end was inserted into the EcoRI site of pBluescriptIISK+ so that the hTRT is oriented in the same direction as the T7 promoter. The selectable marker used with this vector is
25 ampicillin.

pGRN166 (not of the invention)

This vector was constructed for the expression and mutagenesis of TRT sequences in mammalian cells. The coding sequence is operably linked to a T7 promoter. The EcoRI
30 fragment from pGRN151 containing the hTRT ORF with a Kozak front end and IBI flag at the back end was inserted into the EcoRI site of pBluescriptIISK+ so that the

hTRT ORF is oriented in the same direction as the T7 promoter. The insert included full length cDNA of hTRT minus 5' and 3' UTR, FLAG sequence (Immunex Corp, Seattle WA), and Kozak sequences. The selectable marker used with this vector is ampicillin.

5

pGRN167

AvRII-StuI fragment from pGRN144 containing the 5' end of the hTRT ORF was inserted into the XbaI-StuI sites of pBBS161.

10 **pGRN168 (not of the invention)**

The EcoRI fragment from pGRN145 containing the optimized hTRT expression cassette was inserted into the EcoRI site of pIND such that the hTRT coding sequence is in the same orientation as the miniCMV promoter.

15 **pGRN169 (not of the invention)**

The EcoRI fragment from pGRN145 containing the optimized hTRT expression cassette was inserted into the EcoRI site of pIND such that the hTRT is in the reverse orientation from the miniCMV promoter.

20 **pGRN170 (not of the invention)**

The EcoRI fragment from pGRN145 containing the optimized hTRT expression cassette was inserted into the EcoRI site of pIND(sp1) such that the hTRT is in the opposite orientation from the miniCMV promoter.

25 **pGRN171**

The Eco47III-NarI fragment from pGRN163 was inserted into the Eco47III-NarI sites of pGRN167, putting the M1L mutation into a fragment of the hTRT genomic DNA.

pGRN172

30 The BamHI-StuI fragment from pGRN171 containing the Met to Leu mutation in the hTRT ORF was inserted into the BglII-NruI sites of pSEAP2-Basic.

pGRN173

The EcoRV-Eco47III fragment from pGRN144 containing the 5' end of the hTRT promoter region was inserted into the SrfI-Eco47III sites of pGRN172. This makes a
5 promoter reporter plasmid that contains the promoter region of hTRT from approximately 2.3 kb upstream from the start of the hTRT ORF to just after the first intron in the coding region, with the Met1-->Leu mutation.

pGRN174 (not of the invention)

10 The EcoRI fragment from pGRN145 containing the "optimized" hTRT expression cassette was inserted into the EcoRI site of pIND(sp1) such that the hTRT is in the same orientation as the miniCMV promoter.

EXAMPLE 7 (Technical Guidance)

15 RECONSTITUTION OF TELOMERASE ACTIVITY

A. Co-Expression of hTRT and hTR *in vitro*

In this example, the coexpression of hTRT and hTR using an *in vitro* cell-free expression system is described. These results demonstrate that the hTRT polypeptide encoded by pGRN121 encodes a catalytically active telomerase protein and
20 that *in vitro* reconstitution (IVR) of the telomerase RNP can be accomplished using recombinantly expressed hTRT and hTR.

Telomerase activity was reconstituted by adding linearized plasmids of hTRT (pGRN121; 1 µg DNA digested with Xba I) and hTR (phTR+1; 1 µg digested with FspI) to a coupled transcription-translation reticulocyte lysate system (Promega
25 TNTTM). phTR+1 is a plasmid which, when linearized with FspI and then transcribed by T7 RNA polymerase, generates a 445 nucleotide transcript beginning with nucleotide +1 and extending to nucleotide 446 of hTR (Autexier et al., 1996, *EMBO J* 15:5928). For a 50 µl reaction the following components were added: 2 µl TNTTM buffer, 1 µl TNTTM T7 RNA polymerase, 1 µl 1 mM amino acid mixture, 40 units RnasinTM RNase
30 inhibitor, 1 µg each linearized template DNA, and 25 µl TNTTM reticulocyte lysate. Components were added in the ratio recommended by the manufacturer and were

incubated for 90 min at 30°C. Transcription was under the direction of the T7 promoter and could also be carried out prior to the addition of reticulocyte lysate with similar results. After incubation, 5 and 10 µl of the programmed transcription-translation reaction were assayed for telomerase activity by TRAP as previously described (Autexier et al., *supra*) using 20 cycles of PCR to amplify the signal.

The results of the reconstitution are shown in Figure 10. For each transcription/translation reaction assayed there are 3 lanes: The first 2 lanes are duplicate assays and the third lane is a duplicate sample heat denatured (95°C, 5 min) prior to the TRAP phase to rule out PCR generated artifacts.

As shown in Figure 10, reticulocyte lysate alone has no detectable telomerase activity (lane 6). Similarly, no detectable activity is observed when either hTR alone (lane 1) or full length hTERT gene (lane 4) are added to the lysate. When both components are added (lane 2), telomerase activity is generated as demonstrated by the characteristic repeat ladder pattern. When the carboxyl-terminal region of the hTERT gene is removed by digestion of the vector with *NcoI* ("truncated hTERT") telomerase activity is abolished (lane 3). Lane 5 shows that translation of the truncated hTERT alone does not generate telomerase activity. Lane "R8" shows a positive control for a telomerase product ladder generated by TRAP of TSR8, a synthetic telomerase product having a nucleotide sequence of 5'-ATTCCGTCGAGCAGAGTTAG[GGTTAG]₇-3'.

B. Mixing of hTERT and hTR *in vitro*

In vitro reconstitution of telomerase activity was also accomplished by mixing. hTERT was transcribed and translated as described *supra*, but without the addition of the hTR plasmid. Reconstitution of the telomerase RNP was then accomplished by mixing the hTERT translation mixture with hTR (previously generated by T7 RNA polymerase transcription from pHTR+1-Fsp) in the ratio of 2 µl of hTERT translation mix to 2 µl of hTR (1 ug) then incubated for 90 minutes at 30° C. This method of hTERT/hTR reconstitution is referred to as "linked reconstitution" or "linked IVR." Telomerase activity is present (i.e., can be detected) in this mixture. Improved signal was observed following partial purification of the activity by DEAE chromatography. In this case Millipore (RTM) Ultrafree-MC DEAE Centrifugal Filter

Devices were used according to the manufacturer's directions). The buffers used were hypo0.1, hypo0.2, and hypo1.0, where hypo is 20 mM Hepes-KOH, pH 7.9, 2 mM MgCl₂, 1 mM EGTA, 10 % glycerol, 0.1 % NP-40, 1 mM DTT, 1 mM Na-metabisulfite, 1 mM benzamidine, and 0.2 mM phenylmethylsulfonylfluoride (PMSF), and where 0.1, 0.2 and 1.0 refers to 0.1, 0.2 or 1.0 M KCL. The filters were pre-conditioned with hypo1.0, washed with hypo0.1, the reconstituted telomerase was loaded, the column was washed with hypo0.1 then hypo0.2, and the reconstituted telomerase was eluted with hypo1.0 at half the volume as was loaded. This formulation could be stored frozen at -70°C and retains activity.

10 Telomerase activity was assayed in a two step procedure. In step one, a conventional telomerase assay was performed as described in Morin, 1989, *Cell* 59: 521, except no radiolabel was used. In step two, an aliquot was assayed by the TRAP procedure for 20-30 cycles as described *supra*. The conventional assay was performed by assaying 1-10 µl of reconstituted telomerase in 40-50 µl final volume of 25 mM
15 Tris-HCl, pH 8.3, 50 mM K-acetate, 1 mM EGTA, 1 mM MgCl₂, 2 mM dATP, 2 mM TTP, 10 uM dGTP, and 1 uM primer (usually M2, 5'-AATCCGTCGAGCAGAGTT) at 30° C for 60-180 minutes. The reaction was stopped by heating to 95° C for 5 minutes and 1-10 µl of the first step mixture was carried onto the step two TRAP reaction (50 ul).

20 In additional experiments, the synthesis of hTRT and hTR during *in vitro* reconstitution was monitored by ³⁵S-methionine incorporation and Northern blotting, respectively. Proteins of approximately the predicted size were synthesized for hTRT (127 kD), hTRT-Nco (85 kD), and pro90hTRT (90 kD) in approximately equal molar amounts relative to each other. The Northern analysis indicated hTR synthesis was the
25 correct size (445 nucleotides) and predominantly intact.

 Variations of the reconstitution protocols, *supra*, will be apparent to those of skill. For example, the time and temperature of reconstitution, and presence or concentration of components such as monovalent salt (e.g., NaCl, KCl, potassium acetate, potassium glutamate, and the like), divalent salt (MgCl₂, MnCl₂, MgSO₄, and
30 the like), denaturants (urea, formamide, and the like), detergents (NP-40, Tween (RTM), CHAPS, and the like), and alternative improved purification procedures (such

as immunoprecipitation, affinity or standard chromatography) can be employed. These and other parameters can be varied in a systematic way to optimize conditions for particular assays or other reconstitution protocols.

5 C. Reconstitution Using hTERT Variants and Fusion Proteins

Reconstitution of telomerase catalytic activity occurred when EGFP-hTERT, a fusion of the enhanced green fluorescent protein to hTERT (see Examples 6 and 15), or epitope-tagged hTERT (IBI FLAG, see Example 6) both reconstituted telomerase activity at approximately wild-type levels were coexpressed
10 with hTR.

In contrast, telomerase activity was not reconstituted when a variant hTERT, pro90hTERT (missing RT motifs B', C, D, and E) was used. This demonstrates that pro90hTERT does not possess full telomerase catalytic activity, although it may have other partial activities (e.g., RNA [i.e. hTR] binding ability and function as
15 dominant-negative regulator of telomerase *in vivo* as described *supra*).

D. Assay of *in vitro* Reconstituted Telomerase Activity Using the Gel Blot and Conventional Telomerase Assay

The following example demonstrates that *in vitro* reconstituted (IVR)
20 telomerase can be assayed using conventional telomerase assays in addition to amplification-based assays (i.e., TRAP). IVR telomerase as described in part (B), *supra* (the "linked reconstitution method") followed by DEAE purification, as described *supra* was assayed using the gel blot assay using the following reaction conditions; 1-10 µl of linked IVR telomerase in 40 µl final volume of 25 mM Tris-HCl, pH 8.3, 50 mM
25 K-acetate, 1 mM EGTA, 1 mM MgCl₂, 0.8 mM dATP, 0.8 mM TTP, 1.0 mM dGTP, and 1 µM primer (M2, *supra*; or H3.03, 5'-TTAGGGTTAGGGTTAGGG) at 30°C for 180 minutes. The telomeric DNA synthesized was isolated by standard procedures, separated on a 8 % polyacrylamide, 8 M urea gel, transferred to a nylon membrane, and probed using the ³²P-(CCCTAA)_n riboprobe used in the dot-blot assay. The probe
30 identified a six nucleotide ladder in the lane representing 10 µl of IVR telomerase that was equivalent to the ladder observed for 5 µl of native nuclear telomerase purified by

mono Q and heparin chromatography. The results show that IVR telomerase possesses processive telomerase catalytic activity equivalent to native telomerase.

Linked IVR telomerase was also assayed by the conventional ^{32}P -dGTP incorporation telomerase assay. IVR telomerase prepared by the linked reconstitution method followed by DEAE purification, as described above, was assayed under both processive and non-processive reaction conditions. Assay conditions were 5-10 μl of linked IVR telomerase in 40 μl final volume of 25 mM Tris-HCl, pH 8.3, 50 mM K-acetate, 1 mM EGTA, 1 mM MgCl_2 , 2 mM dATP, 2 mM TTP, with 10 μM ^{32}P -dGTP (72 Ci/mmol) [for assay of processive conditions] or 1 μM ^{32}P -dGTP (720 Ci/mmol) [for non-processive], and 1 μM primer (i.e., H3.03, *supra*) at 30°C [for the processive reaction] or 37°C [for the non-processive reaction] for 180 minutes. The telomeric DNA synthesized was isolated by standard procedures and separated on a 8 % polyacrylamide, 8 M urea gel sequencing gel. The processive reaction showed a weak six nucleotide ladder consistent with a processive telomerase reaction, and the non-processive reaction added one repeat, a pattern equivalent to a control reaction with a native telomerase preparation. Conventional assays using IVR telomerase are useful in screens for telomerase modulators, as described herein, as well as other uses such as elucidation of the structural and functional properties of telomerase.

20 E. *In vitro* Reconstituted Telomerase Recognizes Primer 3' Termini

This experiment demonstrates that IVR telomerase recognizes primer 3' termini equivalently to native (purified) telomerase. Telomerase forms a base-paired duplex between the primer 3' end and the template region of hTR and adds the next specified nucleotide (Morin, 1989, *supra*). To verify that IVR (recombinant) telomerase has the same property, the reactions of primers with ---GGG or ---TAG 3' termini (AATCCGTCGAGCAGAGGG and AATCCGTCGAGCAGATAG) were compared to a primer having a ---GTT 3' terminus (M2 *supra*) using IVR and native telomerase assayed by the two step conventional/TRAP assay detailed above. The product ladders of the ---GGG and ---TAG primers were shifted +4 and +2, respectively, when compared to the standard primer (---GTT 3' end), the same effect as was observed with native telomerase. This experiment demonstrates IVR and native telomerases recognize

primer termini in a similar manner.

These results (along with the results *supra* showing that IVR telomerase possesses both processive and non-processive catalytic activity) indicate that IVR telomerase has similar structure and properties compared to native or purified telomerase.

EXAMPLE 8 (Technical Guidance)

PRODUCTION OF ANTI-hTERT ANTIBODIES

A. Production of Anti-hTERT Antibodies Against hTERT Peptides

To produce anti-hTERT antibodies, the following peptides from hTERT were synthesized with the addition of C (cysteine) as the amino terminal residue (see Figure 54).

S-1: FFY VTE TTF QKN RLF FYR KSV WSK

S-2: RQH LKR VQL RDV SEA EVR QHR EA

S-3: ART FRR EKR AER LTS RVK ALF SVL NYE

A-3: PAL LTS RLR FIP KPD GLR PIV NMD YVV

The cysteine moiety was used to immobilize (i.e., covalently link) the peptides to BSA and KLH [keyhole limpet hemocyanin] carrier proteins. The KLH-peptides were used as antigen. The BSA-peptide conjugates served as material for ELISAs for testing the specificity of immune antisera.

The KLH-peptide conjugates were injected into New Zealand White rabbits. The initial injections are made by placing the injectant proximal to the axillary and inguinal lymph nodes. Subsequent injections were made intramuscularly. For initial injections, the antigen was emulsified with Freund's complete adjuvant; for subsequent injections, Freund's incomplete adjuvant was used. Rabbits follow a three week boost cycle, in which 50 ml of blood yielding 20-25 ml of serum is taken 10 days after each boost. Antisera against each of the four peptides recognized the hTERT moiety of recombinant hTERT fusion protein (GST-HIS₈-hTERT-fragment 2426 to 3274); see Example 6) on western blots.

Using a partially purified telomerase fraction from human 293 cells (approximately 1000-fold purification compared to a crude nuclear extract) that was

produced as described in PCT application No. 97/06012 and affinity purified anti-S-2 antibodies, a 130 kd protein doublet could be detected on a western blot. A sensitive chemiluminescence detection method was employed (SuperSignal chemiluminescence substrates, Pierce) but the signal on the blot was weak, suggesting that hTERT is present in low or very low abundance in these immortal cells. The observation of a doublet is consistent with a post-translational modification of hTERT, i.e., phosphorylation or glycosylation.

For affinity purification, the S-2 peptide was immobilized to SulfoLink (Pierce, Rockford IL) through its N-terminal Cysteine residue according to the manufacturer's protocol. First bleed serum from a rabbit immunized with the KLH-S-2 peptide antigen was loaded over a the S-2-SulfoLink and antibodies specifically bound to the S-2 peptide were eluted.

B. Production of Anti-hTERT Antibodies Against hTERT Fusion Proteins

GST-hTERT fusion proteins were expressed in *E. coli* as the GST-hTERT fragment #4 (nucleotides 3272-4177) and the GST-HIS8 -hTERT fragment #3 (nucleotides 2426 to 3274) proteins described in Example 6. The fusion proteins were purified as insoluble protein, and the purity of the antigens was assayed by SDS polyacrylamide gels and estimated to be about 75% pure for the GST-hTERT fragment #4 recombinant protein and more than 75% pure for GST-HIS8 -hTERT fragment #3 recombinant protein. Routine methods may be used to obtain these and other fusion proteins at a purity of greater than 90%. These recombinant proteins were used to immunize both rabbits and mice, as described above.

The first and second bleeds from both the mice and rabbits were tested for the presence of anti-hTERT antibodies after removal of anti-GST antibodies using a matrix containing immobilized GST. The antisera were tested for anti-hTERT antibodies by Western blotting using immobilized recombinant GST-hTERT fusion protein, and by immunoprecipitation using partially purified native telomerase enzyme. While no signal was observed in these early bleeds, titers of anti-hTERT antibodies, as expected, increased in subsequent bleeds.

EXAMPLE 9 (Technical Guidance)

DETECTION OF AN hTRT mRNA CORRESPONDING TO $\Delta 182$ RNA

VARIANT

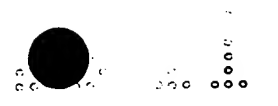
Poly A⁺ RNA from human testis and the 293 cell line was analyzed for hTERT mRNA using RT-PCR and nested primers. The first primer set was TCP1.1 and TCP1.15; the second primer set was TCP1.14 and BTCP6. Amplification from each gave two products differing by 182 bp; the larger and smaller products from testis RNA were sequenced and found to correspond exactly to pGRN121 (Figure 16) and the 712562 clone (Figure 18), respectively. The variant hTERT RNA product has been observed in mRNA from SW39i, OVCAR4, 293, and Testes.

Additional experiments were carried out to demonstrate that the $\Delta 182$ cDNA was not an artifact of reverse transcription. Briefly, full-length hTRT RNA (i.e., without the deletion) was produced by *in vitro* transcription of pGRN121 for use as a template for RT-PCR. Separate cDNA synthesis reactions were carried out using Superscript® reverse transcriptase (Bethesda Research Laboratories, Bethesda MD) at 42° or 50°C, and with random-primers or a specific primer. After 15 PCR cycles the longer product was detectable; however, the smaller product (i.e., corresponding to the deletion) was not detectable even after 30 or more cycles. This indicates that the RT-PCR product is not an artifact.

EXAMPLE 10 (Technical Guidance)

SEQUENCING OF TESTIS hTRT mRNA

The sequence of the testis form of hTRT RNA was determined by direct manual sequencing of DNA fragments generated by PCR from testis cDNA (Marathon Testes cDNA, Clontech, San Diego CA) using a ThermoSequenase radiolabeled terminator cycle sequencing kit (Amersham Life Science). The PCR step was performed by a nested PCR, as shown in Table 8. In all cases a negative control reaction with primers but no cDNA was performed. The absence of product in the



control reaction demonstrated that the products derived from the reaction with cDNA present were not due to contamination of hTERT from pGRN121 or other cell sources (e.g., 293 cells). The DNA fragments were excised from agarose gels to purify the DNA prior to sequencing.

- 5 The testis mRNA sequence corresponding to bases 27 to 3553 of the pGRN121 insert sequence, and containing the entire hTERT ORF (bases 56 to 3451) was obtained. There were no differences between the testis and the pGRN121 sequences in this region.

TABLE 8

Fragment	Primer Set 1	Primer Set 2	Final Size	Primers For Seq
OA	na	K320/K322	208	K320,K322
A	K320/TCP1.43	TCP1.40/TCP1.34	556	TCP1.52,TCP1.39,K322,TCP1.40,TCP1.41,TCP1.30,TCP1.34,TCP1.49
B	TCP1.42/TCP1.32B	TCP1.35/TCP1.21	492	TCP1.35,TCP1.28,TCP1.38,TCP1.21,TCP1.46,TCP1.33,TCP1.48
C	TCP1.65/TCP1.66	TCP1.67/TCP1.68	818	TCP1.67,TCP1.32,TCP1.69,TCP1.68,TCP1.24,TCP1.44,K303
D2	K304/bllTCP6	LT1/TCP1.6	546	LT2,LT1,TCP1.6,bTCP4,TCP1.13,TCP1.77,TCP1.1
D3	TCP1.12/TCP1.7	TCP1.14/TCP1.15	604	TCP1.6,TCP1.14,TCP1.73,TCP1.78,TCP1.25,TCP1.15,TCP1.76
EF	na	TCP1.74/TCP1.7	201	TCP1.74,TCP1.7,TCP1.75,TCP1.15,TCP1.3
E	TCP1.3/TCP1.4	TCP1.2/TCP1.9	687	TCP1.2,TCP1.8,TCP1.9,TCP1.26
F	TCP1.26/UTR2	TCP1.10/TCP1.4	377	TCP1.4,TCP1.10,TCP1.11

EXAMPLE 11 (Technical Guidance)
DETECTION OF hTERT mRNA BY RNASE PROTECTION

5 RNase protection assays can be used to detect, monitor, or diagnose the presence of an hTERT mRNA or variant mRNA. One illustrative RNase protection probe is an *in vitro* synthesized RNA comprised of sequences complementary to hTERT mRNA sequences and additional, non-complementary sequences. The latter sequences are included to distinguish the full-length probe from the fragment of the probe that results from a positive result in the assay: in a positive assay, the complementary
10 sequences of the probe are protected from RNase digestion, because they are hybridized to hTERT mRNA. The non-complementary sequences are digested away from the probe in the presence of RNase and target complementary nucleic acid.

Two RNase protection probes are described for illustrative purposes; either can be used in the assay. The probes differ in their sequences complementary to
15 hTERT, but contain identical non-complementary sequences, in this embodiment, derived from the SV40 late mRNA leader sequence. From 5'-3', one probe is comprised of 33 nucleotides of non-complementary sequence and 194 nucleotides of sequence complementary to hTERT nucleotides 2513 - 2707 for a full length probe size of 227 nucleotides. From 5'-3', the second probe is comprised of 33 nucleotides of
20 non-complementary sequence and 198 nucleotides of sequence complementary to hTERT nucleotides 2837 - 3035 for a full length probe size of 231 nucleotides. To conduct the assay, either probe can be hybridized to RNA, i.e., polyA+ RNA, from a test sample, and T1 ribonuclease and RNase A are then added. After digestion, probe RNA is purified and analyzed by gel electrophoresis. Detection of a 194 nucleotide fragment of
25 the 227 nucleotide probe or a 198 nucleotide fragment of the 231 nucleotide probe is indicative of hTERT mRNA in the sample.

The illustrative RNase protection probes described in this example can be generated by in vitro transcription using T7 RNA polymerase. Radioactive or otherwise labeled ribonucleotides can be included for synthesis of labeled probes. The templates for the *in vitro* transcription reaction to produce the RNA probes are PCR products. These illustrative probes can be synthesized using T7 polymerase following PCR amplification of pGRN121 DNA using primers that span the corresponding complementary region of the hTRT gene or mRNA. In addition, the downstream primer contains T7 RNA polymerase promoter sequences and the non-complementary sequences.

For generation of the first RNase protection probe, the PCR product from the following primer pair (T701 and reverse01) is used:

T701 5'-GGGAGATCT TAATACGACTCACTATAG ATTCA GGCCATGGTG CTGCGCCGGC TGTC A GGCTCCC ACGACGTAGT CCATGTTTAC-3'; and reverse01 5'-GGGTCTAGAT CCGGAAGAGTGT CTGGAGCAAG-3'.

For generation of the second RNase protection probe, the PCR product from the following primer pair (T702 and reverse02) is used:

T702 5'-GGGAGATCT TAATACGACTCACTATAG ATTCA GGCCATGGTG CTGCGCCGGC TGTC A GGGCG GCCTTCTGGA CCACGGCATA CC-3'; and reverse02 5'-G GTCTAGA CGATATCC ACAGGGCCTG GCGC-3'.

EXAMPLE 12 (Technical Guidance)

CONSTRUCTION OF A PHYLOGENETIC TREE COMPARING hTRT AND OTHER REVERSE TRANSCRIPTASES

A phylogenetic tree (Figure 6) was constructed by comparison of the seven RT domains defined by Xiong and Eickbush (1990, *EMBO J.* 9:3353). After sequence alignment of motifs 1, 2, and A-E from 4 TRTs, 67 RTs, and 3 RNA polymerases, the tree was constructed using the NJ (Neighbor Joining) method (Saitou and Nei, 1987, *Mol. Biol. Evol.* 4:406). Elements from the same class that are located on the same branch of the tree are simplified as a box. The length of each box corresponds to the most divergent element within that box.

The TRTs appear to be more closely related to RTs associated with

msDNA, group II introns, and non-LTR (Long Terminal Repeat) retrotransposons than to the LTR-retrotransposon and viral RTs. The relationship of the telomerase RTs to the non-LTR branch of retroelements is intriguing, given that these latter elements have replaced telomerase for telomere maintenance in *Drosophila*. However, the most striking finding is that the TRTs form a discrete subgroup, almost as closely related to the RNA-dependent RNA polymerases of plus-stranded RNA viruses such as poliovirus as to any of the previously known RTs. Considering that the four telomerase genes come from evolutionarily distant organisms -- protozoan, fungi, and mammal -- this separate grouping cannot be explained by lack of phylogenetic diversity in the data set. Instead, this deep bifurcation suggests that the telomerase RTs are an ancient group, perhaps originating with the first eukaryote.

GenBank protein identification or accession numbers used in the phylogenetic analysis were: msDNAs (94535, 134069, 134074, 134075, 134078), group II introns (483039, 101880, 1332208, 1334433, 1334435, 133345, 1353081), mitochondrial plasmid/RTL (903835, 134084), non-LTR retrotransposons (140023, 84806, 103221, 103353, 134083, 435415, 103015, 1335673, 85020, 141475, 106903, 130402, U0551, 903695, 940390, 2055276, L08889), LTR retrotransposons (74599, 85105, 130582, 99712, 83589, 84126, 479443, 224319, 130398, 130583, 1335652, 173088, 226407, 101042, 1078824), hepadnaviruses (I 18876, 1706510, 118894), caulimoviruses (331554, 130600, 130593, 93553), retroviruses (130601, 325465, 74601, 130587, 130671, 130607, 130629, 130589, 130631, 1346746, 130651, 130635, 1780973, 130646). Alignment was analyzed using ClustalW 1.5 [J. D. Thompson, D. G. Higgins, T. J. Gibson, *Nucleic Acids Res.* 22, 4673 (1994)] and PHYLIP 3.5 [J. Felsenstein, *Cladistics* 5, 164 (1989)].

25

EXAMPLE 13 (Technical Guidance)

TRANSFECTION OF CULTURED HUMAN FIBROBLASTS (BJ) WITH CONTROL PLASMID AND PLASMID ENCODING hTERT

This example demonstrates that expression of recombinant hTERT protein in a mammalian cell results in the generation of an active telomerase.

30

Subconfluent BJ fibroblasts were trypsinized and resuspended in fresh

medium (DMEM/199 containing 10% Fetal Calf Serum) at a concentration of 4×10^6 cells/ml. The cells were transfected using electroporation with the BioRad Gene Pulser™ electroporator. Optionally, one may also transfect cells using Superfect™ reagent (Qiagen) in accordance with the manufacturer's instructions. For

5 electroporation, 500 μ l of the cell suspension were placed in an electroporation cuvette (BioRad, 0.4 cm electrode gap). Plasmid DNA (2 μ g) was added to the cuvettes and the suspension was gently mixed and incubated on ice for 5 minutes. The control plasmid (pBBS212) contained no insert behind the MPSV promoter and the experimental plasmid (pGRN133) expressed hTRT from the MPSV promoter.

10 The cells were electroporated at 300 Volts and 960 μ FD. After the pulse was delivered, the cuvettes were placed on ice for approximately 5 minutes prior to plating on 100 mm tissue culture dishes in medium. After 6 hours, the medium was replaced with fresh medium. 72 hours after the transfection, the cells were trypsinized, washed once with PBS, pelleted and stored frozen at -80°C. Cell extracts

15 were prepared at a concentration of 25,000 cells/ μ l by a modified detergent lysis method (see Bodnar et al., 1996, *Exp. Cell Res.* 228:58; Kim et al., 1994, *Science* 266:2011, and as described in patents and publications relating to the TRAP assay, *supra*) and telomerase activity in the cell extracts was determined using a modified PCR-based TRAP assay (Kim et al., 1994, Bodnar et al., 1996). Briefly, 5×10^4 cell

20 equivalents were used in the telomerase primer extension portion of the reaction. While the extract is typically taken directly from the telomerase extension reaction to the PCR amplification, one may also extract once with phenol/chloroform and once with chloroform prior to the PCR amplification. One-fifth of the material was used in the PCR amplification portion of the TRAP reaction (approximately 10,000 cell

25 equivalents). One half of the TRAP reaction was loaded onto the gel for analysis, such that each lane in Figure 25 represents reaction products from 5,000 cell equivalents. Extracts from cells transfected with pGRN133 were positive for telomerase activity while extracts from untransfected (not shown) or control plasmid transfected cells showed no telomerase activity. Similar experiments using RPE cells

30 gave the same result.

Reconstitution in BJ cells was also carried out using other hTERT constructs (i.e., pGRN145, pGRN155 and pGRN138). Reconstitution using these constructs appeared to result in more telomerase activity than in the pGRN133 transfected cells.

5 The highest level of telomerase activity was achieved using pGRN155. As discussed *supra*, pGRN155 is a vector containing the adenovirus major late promoter as a controlling element for the expression of hTERT and was shown to reconstitute telomerase activity when transfected into BJ cells.

10 Notably, when reconstitution using the hTERT-GFP fusion protein pGRN138 (which localizes to the nucleus, see Example 15, *infra*) was performed either *in vitro* (see Example 7) or *in vivo* (transfection into BJ cells) telomerase activity resulted. By transfection into BJ cells, for example, as described *supra*, telomerase activity was comparable to that resulting from reconstitution *in vitro* using pGRN133 or pGRN145.

15 Similar results were obtained upon transfection of normal human retinal pigmented epithelial (RPE) with the hTERT expression vectors of the invention. The senescence of RPE cells is believed to contribute to or cause the disease of age-related macular degeneration. RPE cells treated in accordance with the methods of the invention using the hTERT expression vectors of the invention should exhibit
20 delayed senescence, as compared to untreated cells, and so be useful in transplantation therapies to treat or prevent age-related macular degeneration.

EXAMPLE 14

PROMOTER REPORTER CONSTRUCT

25 This example describes the construction of plasmids in which reporter genes are operably linked to hTERT upstream sequences containing promoter elements. The vectors have numerous uses, including identification of *cis* and *trans* transcriptional regulatory factors *in vivo* and for screening of agents capable of modulating (e.g., activating or inhibiting) hTERT expression (e.g., drug screening). Although a number of
30 reporters may be used (e.g., firefly luciferase, β -glucuronidase, β -galactosidase, chloramphenicol acetyl transferase, and GFP and the like), the human secreted alkaline

phosphatase (SEAP; Clontech) was used for initial experiments. The SEAP reporter gene encodes a truncated form of the placental enzyme which lacks the membrane anchoring domain, thereby allowing the protein to be secreted efficiently from transfected cells. Levels of SEAP activity detected in the culture medium have been
5 shown to be directly proportional to changes in intracellular concentrations of SEAP mRNA and protein (Berger et al., 1988, *Gene* 66:1; Cullen et al., 1992, *Meth. Enzymol.* 216:362).

Four constructs (pGRN148, pGRN150, "pSEAP2 basic" (no promoter sequences = negative control) and "pSEAP2 control" (contains the SV40 early promoter
10 and enhancer) were transfected in triplicate into mortal and immortal cells.

Plasmid pGRN148 was constructed as illustrated in Figure 9. Briefly, a Bgl2-Eco47III fragment from pGRN144 was digested and cloned into the BglII-NruI site of pSeap2Basic (Clontech, San Diego, CA). A second reporter-promoter, plasmid pGRN150, includes sequences from the hTERT intron described in Example 3, to employ
15 regulatory sequences that may be present in the intron. The initiating Met is mutated to Leu, so that the second ATG following the promoter region will be the initiating ATG of the SEAP ORF.

The pGRN148 and pGRN150 constructs (which include the hTERT promoter) were transfected into mortal (BJ cells) and immortal (293) cells. All
20 transfections were done in parallel with two control plasmids: one negative control plasmid (pSEAP basic) and one positive control plasmid (pSEAP control which contains the SV40 early promoter and the SV40 enhancer).

In immortal cells, pGRN148 and pGRN150 constructs appear to drive SEAP expression as efficiently as the pSEAP2 positive control (containing the SV40
25 early promoter and enhancer). In contrast, in mortal cells only the pSEAP2 control gave detectable activity. These results indicate that, as expected, hTERT promoter sequences are active in tumor cells but not in mortal cells.

Similar results were obtained using another normal cell line (RPE, or retinal pigmental epithelial cells). In RPE cells transfected with pGRN150
30 (containing 2.2 KB of upstream genomic sequence), the hTERT promoter region was inactive while the pSEAP2 control plasmid was active.

As noted *supra*, plasmids in which reporter genes are operably linked to hTERT upstream sequences containing promoter elements are extremely useful for identification and screening of telomerase activity modulatory agents, using both transient and stable transfection techniques. In one approach, for example, stable transformants of pGRN148 are made in telomerase negative and telomerase positive cells by cotransfection with a eukaryotic selectable marker (such as *neo*) according to Ausubel et al., 1997, *supra*. The resulting cell lines are used for screening of putative telomerase modulatory agents, for example, by comparing hTERT-promoter-driven expression in the presence and absence of a test compound.

The promoter-reporter (and other) vectors of the invention are also used to identify trans- and cis-acting transcriptional and translational regulatory elements. Examples of cis-acting transcriptional regulatory elements include promoters and enhancers of the telomerase gene. The identification and isolation of cis- and trans-acting regulatory agents provide for further methods and reagents for identifying agents that modulate transcription and translation of telomerase.

EXAMPLE 15 (Technical Guidance)

SUBCELLULAR LOCALIZATION OF hTERT

A fusion protein having hTERT and enhanced green fluorescent protein (EGFP; Cormack et al., 1996, *Gene* 173:33) regions was constructed as described below. The EGFP moiety provides a detectable tag or signal so that the presence or location of the fusion protein can be easily determined. Because EGFP-fusion proteins localize in the correct cellular compartments, this construct may be used to determine the subcellular location of hTERT protein.

25

A. Construction of pGRN138

A vector for expression of an hTERT-EGFP fusion protein in mammalian cells was constructed by placing the EcoRI insert from pGRN124 (see Example 6) into the EcoRI site of pEGFP-C2 (Clontech, San Diego, CA). The amino acid sequence of the fusion protein is provided below. EGFP residues are in bold, residues encoded by the 5' untranslated region of hTERT mRNA are underlined, and the hTERT protein

30

sequence is in normal font.

MVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTCLKFICTTGKLPVPWPPT
LVTTLTLYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEQDTL
VNRIELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQKNGIKVNFKIRHNIEDGSGVQLA
5 DHYQONTPIGDGPVLLPDNHYLSTQSALSQDPNEKRDMVLLFVTAAGITLGMDELYKS
GRTQISSSSFEFAAASTQRCVLLRTWEALAPATPAMPRAPRCRAVRSLLRSHYREVLPLA
TFVRRLLGPQGWRLVQRGDPAAAFRALVAQCLVCPWDARPPPAAPSFRQVSCLELVARVL
QRLCERGAKNVLAFGFALLDGARGGPPEAFTTSVRSYLPNTVTDALRGSGAWGLLLRRVG
10 DDVLVHLLARCALFVLVAPSCAYQVCGPPPLYQLGAATQARPPPHASGPRRRLGCERAWNH
SVREAGVPLGLPAPGARRRGGSASRSLPLPKRPRRGAAPEPERTPVGQGSWAHPGRTRGP
SDRGFCVVSPARPAEEATSLEGALSGTRHSHPSVGRQHHAGPPSTSRPPRPWDTPCPPVY
AETKHFLYSSGDKEQLRPSFLLSSLRPSLTGARRLVETIFLGSRPWMPGTPRRLPRLPQR
YWQMRPLFLELLGNHAQCPYGVLLKTHCPLRAAVTPAAGVCAREKPGQGSVAAPEEEDTDP
15 RRLVQLLRQHSSPWQVYGFVRACLRLVPPGLWGSRRHNERRFLRNTKKFISLGKHAKLSL
QELTWKMSVRDCAWLRRSPGVGCVPAAEHRLREEILAKFLHWLMSVYVVELLRSFFYVTE
TTFQKNRLFFYRPSVWSKLQSIGIRQHLKRVQLRELSEAEVRQHREARPALLTSRLRFIP
KPDGLRPIVNMDYVVGARTFRREKRAERLTSRVKALFVNLNERARRPGLLGASVLGLDD
IHRAWRTFVLVRAQDPPPELYFVKVDVTGAYDTIPQDRLTEVIASIIKPQNTYCVRRYA
20 VVQKAAHGHVRKAFKSHVSTLTDLPYMRQFVAHLQETSPLRDAVVIEQSSSLNEASSGL
FDVFLRFMCHHAVRIRGKSYVQCQGI PQGSILSTLLCSLCYGD MENKLFAGIRRDGLLLR
LVDDFLLVTPHLTHAKTFLRTLVRGVPEYGCVVNLKRTVNVFPVEDEALGGTAFVQMPAH
GLFPWCGLLLDTRTLEVQSDYSSYARTSIRASVTFNRGFKAGRNMRRLFGVLRLLKCHSL
FLDLQVNSLQTVCTNIYKILLQAYRFHACVLQLPFHQVWKNPTFFLRVISDTASLCYS
25 ILKAKNAGMSLGAKGAAGPLPSEAVQWLCHQAFLLKLTRHRVTYVPLLGSLRTAQTLQSR
KLPGTTLTALEAAANPALPSDFKTILD

Other EGFP fusion constructs can be made using partial (e.g., truncated) hTRT coding sequence and used, as described *infra*, to identify activities of particular regions of the hTRT polypeptide.

30

B. Nuclear Localization and Uses of pGRN138

Transfection of NIH 293 and BJ cells with pGRN138 confirmed the nuclear localization of recombinantly expressed hTRT. Cells were transfected with pGRN138 (EGFP-hTRT) and with a control construct (expressing EGFP only).

35 Nuclear localization of the EGFP-hTRT is apparent in both cell types by fluorescence microscopy. As noted *supra*, the pGRN138 hTRT-GFP fusion protein supports reconstitution of telomerase activity in both an *in vitro* transcription translation system and *in vivo* when transfected into BJ cells.

The hTRT-EGFP fusion proteins (or similar detectable fusion proteins)
40 can be used in a variety of applications. For example, the fusion construct described in this example, or a construct of EGFP and a truncated form of hTRT, can be used to assess the ability of hTRT and variants to enter a cell nucleus and/or localize at the

chromosome ends. In addition, cells stably or transiently transfected with pGRN138 are used for screening compounds to identify telomerase modulatory drugs or compounds. Agents that interfere with nuclear localization or telomere localization can be identified as telomerase inhibitors. Tumor cell lines stably expressing EGFP-hTERT can be useful for this purpose. Potential modulators of telomerase will be administered to these transfected cells and the localization of the EGFP-hTERT will be assessed. In addition, FACS or other fluorescence-based methods can be used to select cells expressing hTERT to provide homogeneous populations for drug screening, particularly when transient transfection of cells is employed. In other applications, regions of the hTERT can be mutagenized to identify regions (e.g., residues 193-196 (PRRR) and 235-240 (PKRPRR)) required for nuclear localization, which are targets for anti-telomerase drugs (telomerase activity modulators). Other applications include:

- use of the fusion protein as a fluorescent marker of efficient cell transfection for both transient transfection experiments and when establishing stable cell lines expressing EGFP-hTERT;
- expression of an hTERT-EGFP fusion with mutated nuclear localization signals (deficient for nuclear localization) in immortal cells so that the hTERT mutant-EGFP scavenges all the hTERT of the immortal cells, retaining it in the cytoplasm and preventing telomere maintenance; and
- use as a tagged protein for immunoprecipitation.

EXAMPLE 16 (Technical Guidance)

EFFECT OF MUTATION ON TELOMERASE CATALYTIC ACTIVITY

This example describes hTERT variant proteins having altered amino acids and altered telomerase catalytic activity. Amino acid substitutions followed by functional analysis is a standard means of assessing the importance and function of a polypeptide sequence. This example demonstrates that changes in the reverse transcriptase (RT) and telomerase (T) motifs affect telomerase catalytic activity.

Conventional nomenclature is used to describe mutants: the target residue in the native molecule (hTERT) is identified by one-letter code and position, and the corresponding residue in the mutant protein is indicated by one-letter code. Thus,

for example, "K626A" specifies a mutant in which the lysine at position 626 (i.e., in motif 1) of hTERT is changed to an alanine.

A. Mutation of hTERT FFYxTE Motif

5 In initial experiments, a vector encoding an hTERT mutant protein, "F560A," was produced in which amino acid 560 of hTERT was changed from phenylalanine (F) to alanine (A) by site directed mutagenesis of pGRN121 using standard techniques. This mutation disrupts the TERT FFYxTE motif. The resulting F560A mutant polynucleotide was shown to direct synthesis of a full length hTERT
10 protein as assessed using a cell-free reticulocyte lysate transcription/translation system in the presence of ³⁵S-methionine.

 When the mutant polypeptide was co-translated with hTR, as described in Example 7, no telomerase activity was detected as observed by TRAP using 20 cycles of PCR, while a control hTERT/hTR cotranslation did reconstitute activity. With 30
15 cycles of PCR in the TRAP assay, telomerase activity was observable with the mutant hTERT, but was considerably lower than the control (wild-type) hTERT.

B. Additional Site-Directed Mutagenesis of hTERT Amino Acid Residues

 Conserved amino acids in six RT motifs were changed to alanine using
20 standard site directed mutagenesis techniques (see, e.g., Ausubel, *supra*) to assess their contribution to catalytic activity. The mutants were assayed using IVR telomerase using the two step conventional/TRAP assay detailed in example 7.

 The K626A (motif 1), R631A (motif 2), D712A (motif A), Y717A (motif A), D868A (motif C) mutants had greatly reduced or undetectable telomerase
25 activity, while the Q833A (motif B) and G932A (motif E) mutants exhibited intermediate levels of activity. Two mutations outside the RT motifs, R688A and D897A, had activity equivalent to wild type hTERT. These results were consistent with analogous mutations in reverse transcriptases (Joyce et al., 1994, *Ann. Rev. Biochem.* 63:777) and are similar to results obtained with Est2p (see Lingner, 1997, *Science*
30 276:561). The experiments identify residues in the RT motifs critical and not critical for enzymatic activity and demonstrate that hTERT is the catalytic protein of human

telomerase. The mutations provide variant hTRT polypeptides that have utility, e.g., as dominant/negative regulators of telomerase activity.

Amino acid alignment of the known TRTs identified a telomerase-specific motif, motif T (see *supra*). To determine the catalytic role of this motif in hTRT, a six amino acid deletion in this motif (Δ 560-565; FFYxTE), was
5 constructed using standard site directed mutagenesis techniques (Ausubel, *supra*). The deletion was assayed using IVR telomerase using the two step conventional/TRAP assay detailed in Example 7. The Δ 560-565 mutant had no observable telomerase activity after 25 cycles of PCR whereas wild type hTRT IVR telomerase produced a strong
10 signal. Each amino acid in each residue in motif T was examined independently in a similar manner; mutants F560A, Y562A, T564, and E565A retained intermediate levels of telomerase activity, while a control mutant, F487A, had minimal affect on activity. Notably, mutant F561A had greatly reduced or undetectable telomerase activity, while activity was fully restored in its "revertant", F561A561F. F561A561F changes the
15 mutated position back to its original phenylalanine. This is a control that demonstrates that no other amino acid changes occurred to the plasmid that could account for the decreased activity observed. Thus, the T motif is the first non-RT motif shown to be absolutely required for telomerase activity.

Motif T can be used for identification of TRTs from other organisms and
20 hTRT proteins comprising variants of this motif can be used as a dominant/negative regulator of telomerase activity. Unlike most other RTs, telomerase stably associates with and processively copies a small portion of a single RNA (ie. hTR), thus motif T can be involved in mediating hTR binding, the processivity of the reaction, or other functions unique to the telomerase RT.

25

EXAMPLE 17 (Technical Guidance)

SCREENING FOR TELOMERASE ACTIVITY MODULATORS USING RECOMBINANTLY EXPRESSED TELOMERASE COMPONENTS

This example describes the use of *in vitro* reconstituted telomerase for
30 screening and identifying telomerase activity modulators. The assay described is easily adapted to high-through-put methods (e.g., using multiple well plates and/or robotic

systems). Numerous variations on the steps of the assay will be apparent to one of skill in the art after review of this disclosure.

Recombinant clones for telomerase components (e.g., hTERT and hTR) are transcribed and translated (hTERT only) in an *in vitro* reaction as follows and as described in Example 7 *supra*, using the TNT® T7 Coupled Reticulocyte lysate system (Promega), which is described in U.S. Patent No. 5,324,637, following the manufacturer's instructions:

	<u>Reagent</u>	<u>Amount per reaction (μL)</u>
10	TNT Rabbit Reticulocyte lysate	25
	TNT reaction buffer	2
	TNT T7 RNA Pol.	1
	AA mixture (complete)	1
	Prime RNase inhibitor	1
15	Nuclease-free water	16
	Xba1 cut pGRN121 [hTERT] (0.5 μg)	2
	Fsp1 cut pGRN164 [hTR] (0.5 μg)	2

The reaction is incubated at 30°C for 2 hours. The product is then purified on an ultrafree-MC DEAE filter (Millipore).

The recombinant telomerase product (IVRP) is assayed in the presence and absence of multiple concentrations of test compounds which are solubilized in DMSO (e.g. 10 μM - 100 μM). Test compounds are preincubated in a total volume of 25 μL for 30 minutes at room temperature in the presence of 2.5 μL IVRP, 2.5% DMSO, and 1X TRAP Buffer (20 mM Tris-HCl, pH 8.3, 1.5mM MgCl₂, 63 mM KCl, 0.05%Tween20, 1.0 mM EGTA, 0.1 mg/ml Bovine serum albumin). Following the preincubation, 25 μL of the TRAP assay reaction mixture is added to each sample. The TRAP assay reaction mixture is composed of 1X TRAP buffer, 50μL dNTP, 2.0 μg/ml primer ACX, 4 μg/ml primer U2, 0.8 attomol/ml TSU2, 2 units/50μl Taq polymerase (Perkin Elmer), and 2 μg/ml [³²P]5'end-labeled primer TS (3000Ci/mmol). The reaction tubes are then placed in the PCR thermocycler (MJ Research) and PCR is performed as

follows: 60 min at 30°C, 20 cycles of {30 sec at 94°C, 30 sec. at 60°C, 30 sec. at 72°C}, 1 min at 72°C, cool down to 10°C. The TRAP assay is described, as noted *supra*, in U.S. Patent No. 5,629,154. The primers and substrate used have the sequences: TS Primer (5'-AATCCGTCGAGCAGAGTT-3'); ACX Primer

5 (5'-GCGCGG[CTTACC]3CTAACC-3'); U2 primer

(5'-ATCGCTTCTCGGCCTTTT-3'); TSU2

(5'-AATCCGTCGAGCAGAGTTAAAAGGCCGAGAAGCGAT-3')

After completion of the PCR step, 4 µl of 10X loading buffer containing bromophenol blue is added to each reaction tube and products (20 µl) are run on a
10 12.5% non-denaturing PAGE in 0.5X TBE at 400 V. The completed gel is subsequently dried and the TRAP products are visualized by Phosphorimager or by autoradiography. The telomerase activity in the presence of the test compound is measured by comparing the incorporation of label in reaction product to a parallel reaction lacking the agent.

15

The following clones described in the Examples have been deposited with the American Type Culture Collection, Rockville, MD 20852, USA:

Lambda phage λ 25-1.1 ATCC accession number 209024 deposited 12th May 1997

20 pGRN121 ATCC accession number 209016 deposited 6th May 1997

Lambda phage λGΦ5 ATCC accession number 98505 deposited 14th August 1997

The present invention provides novel methods and materials relating to
25 hTERT and diagnosis and treatment of telomerase-related diseases. While specific examples have been provided, the above description is illustrative and not restrictive.

All publications and patent documents cited in this application are incorporated by reference in their entirety for all purposes to the same extent as if each individual publication or patent document were so individually denoted.

Claims:

1. A recombinant polynucleotide comprising a human telomerase reverse transcriptase (hTERT) promoter sequence.
- 5 2. A polynucleotide as claimed in claim 1, wherein the promoter sequence comprises a sequence of at least 15, optionally at least 50, optionally at least 100, optionally at least 200, or optionally at least 500 nucleotides of Figure 21 or a sequence hybridizable thereto under stringent conditions.
- 10 3. A polynucleotide as claimed in claim 1 or claim 2, wherein the promoter sequence comprises a sequence of at least 15, optionally at least 50, optionally at least at 100, optionally at least 200, or optionally at least 500 nucleotides of nucleotides 1-2440 of Figure 21, or a sequence hybridizable thereto under stringent conditions.
- 15 4. A polynucleotide as claimed in any of claims 1 to 3, wherein the promoter sequence comprises at least the sequence of nucleotides 622 to 2440 of Figure 21 or a sequence hybridizable thereto under stringent conditions.
- 20 5. A polynucleotide as claimed in any of claims 1 to 3, wherein the promoter sequence comprises at least about 15, optionally at least 50, optionally at least 100, optionally at least 200, or optionally at least 500 nucleotides are encoded by lambda phage Gφ5 (ATCC accession no. 98505).
- 25 6. A polynucleotide as claimed in any preceding claim, further comprising a transcribable sequence operably linked to the hTERT promoter sequence.
7. A polynucleotide as claimed in claim 6, wherein the transcribable sequence encodes a protein other than hTERT.

30

8. A polynucleotide as claimed in claim 6 or claim 7, wherein the transcribable sequence causes cell death.
- 5 9. A polynucleotide as claimed in any of claims 6 to 8, wherein the transcribable sequence is a gene encoding a toxin.
10. A polynucleotide as claimed in any of claims 6 to 8, wherein the hTRT promoter sequence is operably linked to a gene encoding protein having an activity that is not itself
10 toxic to a cell, but which renders the cell sensitive to an otherwise nontoxic drug.
11. A polynucleotide as claimed in claim 10, wherein the protein is a Herpes virus thymidine kinase.
- 15 12. A polynucleotide as claimed in any of claims 5 to 11, wherein the hTRT promoter sequence is operably linked to a reporter gene.
13. A polynucleotide as claimed in claim 12, wherein the reporter encodes a protein that is detectable by fluorescence, phosphorescence, or by virtue of possessing an
20 enzymatic activity.
14. A polynucleotide as claimed in claim 12 or claim 13, wherein the detectable protein is firefly luciferase, β -glucuronidase, β -galactosidase, chloramphenicol acetyl transferase, green fluorescent protein, enhanced green fluorescent protein, or the human
25 secreted alkaline phosphatase.
15. An isolated, synthetic, substantially pure, or recombinant polynucleotide having a sequence that is at least about 15 nucleotides in length and comprising a sequence exactly complementary or identical to a contiguous sequence of a nucleic acid encoding
30 the hTRT promoter as set forth in Figure 21 nucleotides 1-2440.

16. A polynucleotide as claimed in claim 15 that is an antisense oligonucleotide.
17. A method of killing a cell in which an endogenous TRT is expressed, comprising introducing a polynucleotide of any of claims 8 to 11 into the cell *in vitro*.
- 5 18. A method of killing a cell, comprising introducing a polynucleotide of claim 10 or claim 11 into the cell *in vitro* and administering gancyclovir.
19. A method as claimed in claim 17 or claim 18, wherein the cell is a human cell.
- 10 20. A method of inhibiting expression of hTRT in a cell, comprising introducing a polynucleotide of claim 15 or claim 16 into the cell *in vitro*.
21. An assay for a compound that modulates hTRT promoter activity, comprising
15 contacting a polynucleotide of any of claims 6 to 14 with the compound and detecting a change in the expression or activity of the transcribable sequence expression product.
22. An assay as claimed in claim 21, wherein expression of the transcribable sequence expression product is detected.
- 20 23. An assay as claimed in claim 21, wherein activity of the transcribable sequence expression product is detected.
24. A method of inactivating an endogenous hTRT promoter in a cell comprising
25 introducing *in vitro* a recombinant polynucleotide capable of recombining with the endogenous hTRT promoter under conditions in which recombination occurs, wherein the recombinant polynucleotide comprises at least about 15, optionally at least 50, optionally at least 100, optionally at least 200, or optionally at least 500 nucleotides of nucleotides 1-2440 of Figure 21.
- 30

25. A polynucleotide of any of claims 1 to 16 for use as a pharmaceutical.
26. The use of a polynucleotide of any of claims 1 to 16 for the manufacture of a
5 medicament for increasing the proliferative capacity of a vertebrate cell, preferably a
mammalian cell.
27. The use of a polynucleotide of any of claims 1 to 16 for the manufacture of a
10 medicament for treating a condition associated with an elevated level of telomerase
activity in a cell.

THIS PAGE BLANK (USPTO)